

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 24, 2004, 10:16:46 ; Search time 6393 Seconds

(without alignments)
11325.786 Million cell updates/sec

Title: US-09-977-418-7

Perfect score: 1987

Sequence: 1 tgaatgaagcctgtccatg.....cnnncttaactgagagaaa 1987

Scoring table: IDENTITY NUC

Searched: 32822875 segs, 18219855908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_est1.*
2: gb_est2.*
3: gb_hrc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	819.2	41.2	878	5	EX402471
2	818	41.2	1139	4	BM926609
3	804.4	40.5	901	7	CP552903
4	742.6	37.4	848	7	CP552903
5	643.2	32.4	690	5	BM006459
6	642.2	32.3	974	5	BE250170
7	638.8	32.1	726	5	EX390404
8	624.6	31.4	726	5	BE792513
9	613.8	30.9	681	6	CA424992
10	613.8	30.9	681	5	BM996902
11	613.2	30.9	676	5	BU624626
12	593	29.8	636	2	BE306561
13	587.4	29.6	654	4	BM720799
14	587.4	29.6	956	2	BE203719
15	585	29.4	966	2	BE203383
16	566	28.5	607	5	BM985044
17	560	28.2	594	6	CA424154
18	557.8	28.1	595	1	AT479356
19	538	27.1	562	4	BT113724
20	535	26.9	553	1	AT079876
21	530.4	26.7	647	5	BO637483
22	527.2	26.5	533	1	AT709377
23	520.8	26.2	937	5	BU178934
24	520.4	26.2	1118	4	BM547264

25	520	26.2	824	7	CP552730
26	519	26.1 <td>550</td> <td>1</td> <td>AJ708019</td>	550	1	AJ708019
27	511	25.7 <td>534</td> <td>1</td> <td>AJ709026</td>	534	1	AJ709026
28	503.4	25.3 <td>624</td> <td>2</td> <td>BE250457</td>	624	2	BE250457
29	502.6	25.3 <td>605</td> <td>2</td> <td>AM844208</td>	605	2	AM844208
30	500.8	25.2 <td>531</td> <td>1</td> <td>AJ709551</td>	531	1	AJ709551
31	500.4	25.2 <td>528</td> <td>1</td> <td>AJ711177</td>	528	1	AJ711177
32	492.6	24.8 <td>951</td> <td>2</td> <td>BE306631</td>	951	2	BE306631
33	489.8	24.7 <td>518</td> <td>1</td> <td>AI1334158</td>	518	1	AI1334158
34	483	24.3 <td>533</td> <td>1</td> <td>AJ709557</td>	533	1	AJ709557
35	480.4	24.2 <td>508</td> <td>6</td> <td>CB267862</td>	508	6	CB267862
36	478.4	24.1 <td>495</td> <td>1</td> <td>AI041115</td>	495	1	AI041115
37	475.8	23.9 <td>506</td> <td>1</td> <td>AJ707934</td>	506	1	AJ707934
38	475.4	23.9 <td>478</td> <td>1</td> <td>MA688131</td>	478	1	MA688131
39	475.2	23.9 <td>481</td> <td>1</td> <td>AI1034359</td>	481	1	AI1034359
40	474	23.9 <td>528</td> <td>1</td> <td>AJ711705</td>	528	1	AJ711705
41	473.4	23.8 <td>505</td> <td>2</td> <td>AM082736</td>	505	2	AM082736
42	473.2	23.8 <td>532</td> <td>1</td> <td>AJ709542</td>	532	1	AJ709542
43	471	23.7 <td>488</td> <td>1</td> <td>AI133328</td>	488	1	AI133328
44	470	23.7 <td>833</td> <td>7</td> <td>CP552429</td>	833	7	CP552429
45	465.2	23.4 <td>518</td> <td>6</td> <td>CD625679</td>	518	6	CD625679

ALIGNMENTS

RESULT 1
LOCUS BX402471/c 878 bp mRNA linear EST 29-APR-2004
DEFINITION BX402471 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
ACCESSION BX402471
VERSION BX402471.2 GI:46874966
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30618825.
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen. This sequence belongs to sequence cluster 545_r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?e=CS1A10102F0VNP1fc=545_r.
Location/Qualifiers

FEATURES
source
1..878
/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="CSOD1039v13"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 41.2%; Score 819.2; DB 5; Length 878;
Best Local Similarity 96.9%; Pred. No. 6e-211;
Matches 854; Conservative 1; Mismatches 22; Indels 4; Gaps 2;

QY	1013	GGGTTTTTCAAATACAGCTGGGTTCGTGAAAAATCAGATGAAACCTCTGCGGCTCGGCG	1072
Db	878	GGGTTTTTCCAATACAGCTGCTGTCTGATTAACATGAAACTCTGCGCTCGCGCC	819
QY	1073	TGCTGGGGCTTCCAGGCAAGGCAAGTGGGGTTGGGGGTGGGGCTGTCTCTTCTCTCC	1132
Db	818	TGCTGGGGGCTTCCAGGCAAGGCAAGTGGGGTTGGGGGTGGGGCTGTCTCTCTCC	759
QY	1133	CACAGGCTGTGTTGNTGGGGCTGCTCCCATGACAGAGATATCATTAACAGATGGA	1192
Db	758	CACAGGCTGTGTTGNTGGGGCTGCTCCCATGACAGAGATATCATTAACAGATGGA	699
QY	1193	GCCAGGGCATGATATGGGGCTTTGGGTCTCTCAGGTTGACCCCACTTCTTGCCACTTC	1253
Db	698	GCCAGGGCATGATATGGGGCTTTGGGTCTCTCAGGTTGACCCCACTTCTTGCCACTTC	639
QY	1253	CCCTCCGGCAGTCAAGCTCTCCATCCATCCCTCTTTATCTATGATATCTATAGGCTCG	1312
Db	638	CCCTCCGGGAGTCAAGCTCTCCATCCATCCCTCTTTATCTATGATATCTATAGGCTCGT	579
QY	1313	TGTGTGTAAACAACACCCCTATGCTGTTCCTTCAATACAGATTAACAATTGTTGTA	1372
Db	578	GTTGTGTAAACAACAACCCCTATGCTGTTCCTTCAATACAGATTAACAATTGTTGTA	519
QY	1373	GGCCAAATTACAGGCTTTCTCAATACAGATTTACATCTCCATTTTCAATTAACGGGGAA	1432
Db	518	GGCCAAATTACAGGCTTTCTCAATACAGATTTACATCTCCATTTTCAATTAACGGGGAA	459
QY	1433	CATCCCCAGAGCACTGAGTGCCTGTGCTTGTACAGAGGTAGATCGAAACCAAGGCTG	1492
Db	458	CATCCCCAGAGCACTGAGTGCCTGTGCTTGTACAGAGGTAGATCGAAACCAAGGCTG	399
QY	1493	TC AACNCTGCTTCAACTCCCACTCTGGGCACTGAGAGTATTTCCCTCAATCTAC	1552
Db	398	TC AACNCTGCTTCAACTCCCACTCTGGGCACTGAGAGTATTTCCCTCAATCTAC	339
QY	1553	CTCTTAAGGCTATGACCCCTCCCACTCTTCAGCTGGGGGATGGGGGAGATCAAG	1612
Db	338	CTCTTAAGGCTATGACCCCTCCCACTCTTCAGCTGGGGGATGGGGGAGATCAAG	279
QY	1613	GA AAAAGCCCCATCTCCATCTGGGATGGAACCTTCATCACTAAGCCTTAACCTTGGAAT	1672
Db	278	GA AAAAGCCCCATCTCCATCTGGGATGGAACCTTCATCACTAAGCCTTAACCTTGGAAT	219
QY	1673	GCTGCTGCCCCCACTGACTCTTGTTTCGTCTCCCACTACAGAGAGGGGTGGAAGGG	1732
Db	218	GCTGCTGCCCCCACTGACTCTTGTTTCGTCTCCCACTACAGAGAGGGGTGGAAGGG	159
QY	1733	AAGGCTGGGTCTCACTTACAGAGGGTCCCAAGGGCAAGTCAAGCTCTCTCTCCATGCTT	1792
Db	158	AAGGCTGGGTCTCACTTACAGAGGGTCCCAAGGGCAAGTCAAGCTCTCTCTCCATGCTT	99
QY	1793	CTTGCTAGTGTGCTTAAAGGTGGCTCTCACTCCCACTGAGG--CCTTGGGGGA	1851
Db	98	CTTGCTAGTGTGCTTAAAGGTGGCTCTCACTCCCACTGAGG--CCTTGGGGGA	39
QY	1852	GGACTGGGGAAGGGGCGCTGGGAGAGACCTTACGCTGGAAC	1892
Db	38	GGACTGGGGAAGGGGCG--TGGAAGACCTTACGCTGGAAC	1

RESULT 2	BM926609	1139 bp	mRNA	linear	EST 12-MAR-2002
LOCUS	BM926609				
DEFINITION	BM926609	1139 bp	mRNA	linear	EST 12-MAR-2002
	ACENECOUT 6644689 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5767011				
	5', mRNA sequence.				
ACCESSION	BM926609				
VERSION	BM926609.1	GI:15376988			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1139)	NIH-MGC	http://mgc.nci.nih.gov/	Unpublished (1999)	Contact: Robert Strauberg, Ph.D.

Email: csagab@remail.lin.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.liml.gov>
Plate: LRAM12824 Row: J Column: 10
High quality sequence: stop: 717.

```

FEATURES
Source
1..1139
Location/Qualifiers
1..1139
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5767017"
/lab_host="RDH10B"
/clone_id="NIH_MGC_122"
note="Organ: pooled lung and spleen; Vector: pCMV-Sport6
Site_1: Nci1; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

```

Query Match	41.2%	Score 818;	DB 4;	Length 1139;
Best Local Similarity	93.3%	Pred. No. 1,4E-210;		
Matches 941;	Conservative	0;	Mismatches 55;	Indels 13; Gaps 8;
QY 605	ACTTAATCCAGATACAGAGGCTGGGCTCAGGAGAGAAAGATCTAAAGAACCCACTGTGG	664		
Db 12	ACTTAATCCAGAGTACAGAGGCTGGGCTCAGGAGAGAAAGATCTAAAGAACCCACTGTGG	71		
QY 665	GTCCAGGGGAATGGGACCGACAGGACATATGGGCAAGCTGTGCAGGACACAGACAGACACA	724		
Db 72	GTCCAGGGGAATGGGACCGACAGGACATATGGGCAAGCTGTGCAGGACACAGACAGACACA	131		
QY 725	AACCCCTGATCTATGAAAGTCTTCGACGGGCAAGGGGACCGAGGACCTTGAAACCTCTTGG	784		
Db 132	AACCCCTGATCTATGAAAGTCTTCGACGGGCAAGGGGACCGAGGACCTTGAAACCTCTTGG	191		
QY 785	GCCAAAGGGGAGTGGGAGAGACAGAGGGGAAGTCAACAGGCAAGGGTGCCTATTAAAGTGA	844		
Db 192	GCCAAAGGGGAGTGGGAGAGACAGAGGGGAAGTCAACAGGCAAGGGTGCCTATTAAAGTGA	251		
QY 845	ACTTAATTGGCCGAGGAGCTCAGCAAGGCGCAAGAGAGACACCGTAGCGGTAAACTTCCCC	904		
Db 252	ACTTAATTGGCCGAGGAGCTCAGCAAGGCGCAAGAGAGAGACACCGTAGCGGTAAACTTCCCC	311		
QY 905	TTTACCAAGCTTCCAAGCCCAACGCGACGAGCAGGCTGCTGGCCAACCCCGTGGCCCAAG	964		
Db 312	TTTACCAAGCTTCCAAGCCCAACGCGACGAGCAGGCTGCTGGCCAACCCCGTGGCCCAAG	371		
QY 965	CCAGCTGCTGTGCGCAGGCGCAGAGCCATGCGCATCTGTATATATGATGGGGTTTTTCCAA	1022		
Db 372	CCAGCTGCTGTGCGCAGGCGCAGAGCCATGCGCATCTGTATATATGATGGGGTTTTTCCAA	431		
QY 1025	TACAGCTGGTTCGTGAAGAAAACGTGATTAACACTCTGCTCCGTCTGCGCTGCTGGGGCTTC	1084		
Db 432	TACAGCTGGTTCGTGAAGAAAACGTGATTAACACTCTGCTCCGTCTGCGCTGCTGGGGCTTC	491		
QY 1085	CAGGCAAGGCCAAGTGGGGTTGGGGGCTGGTCTTCTTCCTTCCACAGGCTGTG	1144		

Db 492 CAGGCAAGGCCACGCTGGGAGTGGGGGCTGCTCTTCTCCCTCCACAGGCGCTGTG 551
QY 1145 TTGNTGGGGGCTGCTCCCATGACAGACAGATCACCTAACAGAGATGAGCGAGGCGATG 1204
Db 552 TTTCTGGGGCTGCTCCCATGACAGACAGATCACCTAACAGAGATGAGCGAGGCGATG 611
QY 1205 ATGGGGCTTGGTCTCGAGGTGAGACCCAGCTTCTGCGACCTTCCCTCC -GGCAG 1263
Db 612 ATGGGGCTTGGTCTCGAGGTGAGACCCAGCTTCTGCGACCTTCCCTCC -GGCAG 671
QY 1264 TCAGCTCTCATCATCCCTCTTAACTAGATCTAAGGCTCGGT -GGTGTAC 1322
Db 672 TCAGCTCTCATCATCCCTCTTAACTAGATCTAAGGCTCGGTGGGTGTAC 731
QY 1323 A-ACACACCCCTATCGTGTCTTCAAAATACAGCATTA -CCATTGCTGAGCCCAAT 1380
Db 732 ACACACACCCCTATCGTGTCTTCAAAATACAGCATTA -CCATTGCTGAGCCCAAT 791
QY 1381 TCAGAGCTTCTCAATCAGATTTTCAATCTTCAATTTTCAACGGGAAATCTCCCG 1440
Db 792 TCAGAGCTTCTCAATCAGATTTTCAATCTTCAATTTTCAACGGGAAATCTCCCG 851
QY 1441 AGCAGCTGAGTGTCTGCTTGTCTGAC -TGAAGTATGATCTGAA - -CCAGGGGTGCA 1495
Db 852 AGCAGCTGAGTGTCTGCTTGTCTGAC -TGAAGTATGATCTGAA - -CCAGGGGTGCA 911
QY 1496 ACNGCTGCTCTCACTCCCACTCTGGGCA -CTGAGAGATTTTCCCTCATTC -TA 1551
Db 912 CAGCTGCTCTCACTCCCACTCTGGGCACTGAGAGAGATTTTCCCTCATCTAA 971
QY 1552 CCTCTCAAGGCTATGACACCCCTCCCAAGCTTTCAGCTGGGGAGTGG 1600
Db 972 CCTCTCAAGGCTATGACACCCCTCCCAAGCTTTCAGAGATTTGGGGG 1020

RESULT 3
CF552903 901 bp mRNA linear EST 22-SEP-2003
LOCUS AGENCOURT 15596050 NIH_MGC_183 Homo sapiens cDNA clone
DEFINITION IMAGE:30529648 5', mRNA sequence.
ACCESSION CF552903
VERSION CF552903.1 GI:34889737
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: gcaps-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDAM617 row: 1 column: 17
High quality sequence stop: 674.
Location/Qualifiers

FEATURES
Source

1..901
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30529648"
/lab_host="DH10B-Tona (T1 and T5 phage resistant)"
/clone_lib="NIH_MGC_183"

ORIGIN
/note="Organ: Pooled muscle (cardiac and skeletal);
Vector: pCMV-Sport6.1; Site 1: EcoRV (destroyed); Site 2:
NotI; Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.7. Library was constructed by Invitrogen."

Query Match 40.5%; Score 804.4; DB 7; Length 901;
Best Local Similarity 97.1%; Pred. No. 6.3e-207;
Matches 872; Conservative 0; Mismatches 18; Indels 8; Gaps 5;

QY 760 GACACAGGAGCTGGAACCTCTTGGCCAGGGAGTGGAGAGAGAGAGTGTAC 819
Db 1 GACACAGGAGCTGGAACCTCTTGG -CAAGGGAGTGGAGAGAGAGAGTGTAC 59
QY 820 AGGCAAGGAGCTGATCTAAGTGAATAATTTGCCAGAGGCTCAGCAAGGCCAAGAGA 879
Db 60 AGGCAAGGAGCTGATCTAAGTGAATAATTTGCCAGAGGCTCAGCAAGGCCAAGAGA 119
QY 880 GACAGCGGTGACGTAACTTCCCTCTTACCAAGCTTCAAGCCCAAGCCAGAGAGAG 939
Db 120 GACAGCGGTGACGTAACTTCCCTCTTACCAAGCTTCAAGCCCAAGCCAGAGAGAG 179
QY 940 CTGCTGACCCAGCCCGTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999
Db 180 CTGCTGACCCAGCCCGTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239
QY 1000 CTGATATATAGTGGGCTTTTTCATATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059
Db 240 CTGATATATAGTGGGCTTTTTCATATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
QY 1060 GCGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
Db 300 GCGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
QY 1120 TCTTCTCTCTCTCCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1179
Db 360 TCTTCTCTCTCTCCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
QY 1180 AACAGAGTGAAGGAGGAGGAGTGGGCTTGGGCTCTCGAGGTTGAGACCCCACT 1239
Db 420 AACAGAGTGAAGGAGGAGGAGTGGGCTTGGGCTCTCGAGGTTGAGACCCCACT 479
QY 1240 TCTTCCCACTTCCCTCC -GGCAGTCACTCTCCATCATCCCTCTTTAATCTAGA 1298
Db 480 TCTTCCCACTTCCCTCC -GGCAGTCACTCTCCATCATCCCTCTTTAATCTAGA 539
QY 1299 ACTTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1357
Db 540 ACTTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
QY 1358 ATTACCATTTGTTGAGGCGCAATTGAGGCTTCTCAATCAGATTTCAATCTCATTT 1417
Db 600 ATTACCATTTGTTGAGGCGCAATTGAGGCTTCTCAATCAGATTTCAATCTCATTT 659
QY 1418 TCATTAGGAGGAGAAATCCCGAGCACTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
Db 660 TCATTAGGAGGAGAAATCCCGAGCACTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
QY 1478 TCTGAACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537
Db 720 TCTGAACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
QY 1538 TTTCCCTCATTTTACCTCTTAAAGCTATGACCCCTCCCAAGCTTTTCCAGCTGGGG - 1595
Db 780 TTTCCCTCATTTTACCTCTTAAAGCTATGACCCCTCCCAAGCTTTTCCAGCTGGGG 839
QY 1596 - -GATGGGGAGGTATAGAGAAAGCCCTC -ATCTCCCATCTGGGATAGGAGCTTCC 1650
Db 840 AAGGGGGGGGATCATAGAGAAAGCCCTCTATTTTCCATCTGGGAGAGAGAACTTC 897

RESULT 4

CEFS52003	LOCUS	848 bp	mRNA	linear	EST 22-SEP-2003
DEFINITION	AGENCOURT_15595803 NIH_MGC_183 Homo sapiens cDNA clone IMAGE:3052976 5', mRNA sequence.				
ACCESSION	CEFS52003				
VERSION	CEFS52003.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics / NIH National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgsabds@imail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: NDM618 row: 9 column: 09 High quality sequence start: 2 High quality sequence stop: 666. Location/Qualifiers 1..848 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3052976" /lab_host="DH10B-Tona (T1 and T5 phage resistant)" /clone_id="NIH MGC 183" /note="Organ: Pooled muscle (cardiac and skeletal); Vector: pCMV-Sport6.1; Site 1: EcoRV (destroyed); Site 2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7. Library was constructed by Invitrogen."				
ORIGIN					
Query Match	37.4%; Score 742.6; DB 7; Length 848;				
Best Local Similarity	96.2%; Pred. No. 3.9e-190;				
Matches	816; Conservative 0; Mismatches 20; Indels 12; Gaps 5;				
QY	516 CCAACTAGGGCCCTCCACTATTCACGAGGCGCAGGAGCTCCCTGGCTGACATG 575				
Db	1 CCCAACTAGGGCCCTCCACTATTCACGAGGCGCAGGAGCTCCCTGGCTGACATG 60				
QY	576 ATGCCCAAGTTTCAGATTTGGCCCTCCGCTCATTTATCCAGATCCAGGGGCTGGGGTCAG 635				
Db	61 ATGCCCAAGTTTCAGATTTGGCCCTCCGCTCATTTATCCAGATCCAGGGGCTGGGGTCAG 120				
QY	636 GGAAGAAGATCTTAAAGAACCCCACTGTGGTCAAGGGGAATGGACACAGACATATGG 695				
Db	121 GGAAGAAGATCTTAAAGAACCCCACTGTGGTCAAGGGGAATGGACACAGACATATGG 180				
QY	696 GCAAGCTCTGAGGACAGACAGACAGACAAACCTCTGATTTAGTAAGTCTTTCAGAGGC 755				
Db	181 GCAAGCTCTGAGGACAGACAGACAGACAAACCTCTGATTTAGTAAGTCTTTCAGAGGC 240				
QY	756 AAGGGGACCAAGGACCTGGAACCTCTTGGCCAAAGGGGAGTGGGAGACAGAGGAGG 815				
Db	241 AAGGGGACCAAGGACCTGGAACCTCTTGGCCAAAGGGGAGTGGGAGACAGAGGAGG 300				
QY	816 TCACAGGCAAGGGTCTTATTTAGTGAATTAATTCGCGAGGGCTTCAGAACGCCAAG 875				
Db	301 TCACAGGCAAGGGTCTTATTTAGTGAATTAATTCGCGAGGGCTTCAGAACGCCAAG 360				

QY	876	GGGAGACAGCCGTGACGGTAAACTTCCCTCTACAGGCTCCAGGCCACGGCCAGCGAG	935	
Db	361	AGGAGACAGCCGTGACGGTAACTTCCCTCTACAGGCTCCAGGCCACGGCCAGCGAG	420	
QY	936	CAGGCTGCTCCGCCACCCCGTGCCTCCCGCCAGCGAGCTGTGTGCCAGGCGAGGCCATGCC	995	
Db	421	CAGGCTGCTCCGCCACCCCGTGCCTCCCGCCAGCGAGCTGTGTGCCAGGCGAGGCCATGCC	480	
QY	996	ACATCTGATATTAATATGGGGTTTTTCCAAATCAGCTGGTTCTGTAAAAAATCGATGAAC	1055	
Db	461	ACATCTGATATTAATATGGGGTTTTTCCAAATCAGCTGGTTCTGTAAAAAATCGATGAAC	540	
QY	1056	TCCTGCGCCCTCCGCGCTGTGGGCGCTCCAGGCGAAGGCAATGGGGTTT-GGGGGTGGG	1114	
Db	541	TCCTGCGCGCTCCGCGCTGTGGGCGCTCCAGGCGAAGGCAATGGGGTTTGGGGGTGGG	600	
QY	1115	GCTGGTCTTCTTCCCTCCCAAGCGCTGTGTCTTCTGGGGCTGCTCCATGACAGACGAT	1174	
Db	601	GCTGGTCTTCTTCCCTCCCAAGCGCTGTGTCTTCTGGGGCTGCTCCATGACAGACGAT	660	
QY	1175	CACCTTAACAGATGAGAGCAGGCGCATGATGGAG--CTTGGGTCCTCGAGTTGGAC	1232	
Db	661	CACCTTAACAGATGAGAGCAGGCGCATGATGGAGGCTTTGGGGTCTCGAGTTGGAC	720	
QY	1233	CCGAGCTTCTTCCGACCTTCCCC--TCGCGAGTCAAGCTTCATTCATCCCTCTTTTA	1290	
Db	721	CCGAGCTTCTTCCGACCTTCCCCCTCGGCGAGTCAGCTTCATTCATCCCTCTTTTA	780	
QY	1291	ATCATGATATTAATATAGGCTC-----GGGTGTGTAACACACACACCCCATC--GTGTCC	1343	
Db	781	ATCATGATATTAATATAGGCTCCTCGGCGGTGTGTAAACAAACACACCCCATCCTTTGTCC	840	
QY	1344	TTCAATA 1351		
Db	841	TTCAATA 848		
RESULT 5	BQ006459/c	690 bp	linear	EST 26-MAR-2002
LOCUS	UI-H-E11-aza-k-12-O-UI.s1	NCI CGAP E11	Homo sapiens	CDNA clone
DEFINITION	IMAGE:5846147.3	, mRNA sequence.		
ACCESSION	BQ006459			
KEYWORDS	BQ006459.1	GI:19731359	EST.	
ORGANISM	Homo sapiens (human)			
SOURCE	Homo sapiens			
REFERENCE	Enkaryotia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 690)			
JOURNAL	NCI-CGAP htcp://www.ncbi.nlm.nih.gov/ncicgap.			
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: sgabbs-rc@mail.nih.gov			
	Tissue Procurement: Dr. Jose Mercuende			
	CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa			
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa			
	Clone Sequencing by: Dr. M. Bento Soares, University of Iowa			
	Clone Distribution: Clone distribution information can be found			
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	Seq primer: M13 FORWARD			
	POLYA=yes			
FEATURES	Location/Qualifiers			
Source	1..690			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:5846147"			
	/tissue_type="Chondrosarcoma"			
	/dev_stage="Adult"			
	/lab_host="DH10B (Life Technologies)"			

/clone_lib="NCI_CGAP_E11"
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
 with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP E11 is a normalized cDNA library containing the
 following tissue(s): Chondrosarcoma. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AACTTGCAC.
 TAG_TISSUE=chondrosarcoma
 TAG_LIB=UI-H-E11
 TAG_SEQ=AACTTGCAC"

ORIGIN

Query Match 32.4%; Score 643.2; DB 5; Length 690;
 Best Local Similarity 99.0%; Pred. No. 3,8e-163;
 Matches 667; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

1252 CCCCTCCGCGAGTCAGCTCCATCCATCCCTCTTATATATATATATATAGGCTCG 1311
 689 CCTCCGCGAGTCAGCTCCATCCATCCCTCTTATATATATATATATAGGCTCG 630

1312 GTGTGTGTAAACA-ACAACCCCTATCGTTGTCCTTCAAAATCTACACATTACATTGTT 1370
 629 GTGTGTGTAAACAACAACCCCTATCGTTGTCCTTCAAAATCTACACATTACATTGTT 570

1371 GAGGCCAAATTCAGAGTTTCTCAATCAATTAATTAATTCATTTTCAATTAAGGGGA 1430
 569 GAGGCCAAATTCAGAGTTTCTCAATCAATTAATTAATTCATTTTCAATTAAGGGGA 510

1431 AATATCCCGAGCAGCTAGTGTGCTTGTCACTGAAGTTAGATCGAACCCAGG 1490
 509 AATATCCCGAGCAGCTAGTGTGCTTGTCACTGAAGTTAGATCGAACCCAGG 450

1491 TGTCAACGCTGCTCTCACTCCCACTCTGAGCACTGAGAGTATTTCCCTCATTTCT 1550
 449 TGTCAACGCTGCTCTCACTCCCACTCTGAGCACTGAGAGTATTTCCCTCATTTCT 390

1551 ACCCTCTAAGGCTATGACACCCCTCCCACTCTTCCAGCTGGGGGATGGGGGAGTCAT 1610
 389 ACCCTCTAAGGCTATGACACCCCTCCCACTCTTCCAGCTGGGGGATGGGGGAGTCAT 330

1611 AGGAAAAGCCCCCATCTCCCATCTGGATAGGAGACCTTCATCAAGCTTAAACCTTGGGA 1670
 329 AGGAAAAGCCCCCATCTCCCATCTGGATAGGAGACCTTCATCAAGCTTAAACCTTGGGA 270

1671 ATGCTCTGCTGCCCCCATCTCTTGGTTCTTCTCCCATCAAGAGAGAGGATGGAGG 1730
 269 ATGCTCTGCTGCCCCCATCTCTTGGTTCTTCTCCCATCAAGAGAGAGGATGGAGG 210

1731 GGAAGGCTGGGCTCTCAATTAGCAGGGGTCCCAAGGGCAAGTCAAGCTCTTCCCTCATGTC 1790
 209 GGAAGGCTGGGCTCTCAATTAGCAGGGGTCCCAAGGGCAAGTCAAGCTCTTCCCTCATGTC 150

1791 CTCTCTGCTAGTGTGCTTAAAGGTGAGCTCTCACTCCCACTCTGGG-CCCTTGGG 1849
 149 CTCTCTGCTAGTGTGCTTAAAGGTGAGCTCTCACTCCCACTCTGGG-CCCTTGGG 90

1850 GAGAGCTGGGAGGGGGCTGGGAGAGACCTGACGTGAACCTGTATACAAATTAAG 1909
 89 GAGAGCTGGGAGGGGGCTGGGAGAGACCTGACGTGAACCTGTATACAAATTAAG 30

1910 GACAGTCTCACAGA 1923
 29 GACAGTCTCACAGA 16

RESULT 6
 BE250170 974 bp mRNA linear EST 13-JUN-2000
 LOCUS 600943555F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960076 5',
 DEFINITION mRNA sequence.
 ACCESSION BE250170.1 GI:9120277
 VERSION BE250170
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 974)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: L1CM52 row: n column: 13
 High quality sequence stop: 692.
 Location/Qualifiers
 1..974
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2960076"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="MDH10B (phage-resistant)"
 /clone_lib="NIH_MGC_17"
 /note="Organ: muscle; Vector: pCMV7; Site 1: EcoRI;
 Site 2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoR/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES

ORIGIN

Query Match 32.3%; Score 642.2; DB 2; Length 974;
 Best Local Similarity 96.0%; Pred. No. 7.7e-163;
 Matches 700; Conservative 0; Mismatches 25; Indels 4; Gaps 4;

782 TTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGGTCTTATTAAGT 841
 2 TTGGCCAAAGGGAGTGGGAGAGAGGAGGAAAGTCAAGGCAAGGGTCTTATTAAGT 61

842 GGAACCTAATTGGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGGTAAACTTC 901
 62 GGAACCTAATTGGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGGTAAACTTC 121

902 CCTCTACAGAGCTTCAAGGCCCAAGGCAAGGAGGCTGCTCCCAACCCGCTGCC 961
 122 CCTCTACAGAGCTTCAAGGCCCAAGGCAAGGAGGCTGCTCCCAACCCGCTGCC 181

962 CAGCCAGCTGGTGGCCAGGCGAGGCAAGCCATGCACTGTATATAGTGGGTTTTC 1021
 182 CAGCCAGCTGGTGGCCAGGCGAGGCAAGCCATGCACTGTATATAGTGGGTTTTC 241

1022 CATAACAGCTGGTGGTGAATAAGTCAATGAATCTGCGCTCTGCGCTGCTGGG 1081
 242 CATAACAGCTGGTGGTGAATAAGTCAATGAATCTGCGCTCTGCGCTGCTGGG 301

1082 CTCACAGGAGGCAAGTGGGGTTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 1141
 302 CTCACAGGAGGCAAGTGGGGTTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 361

ACCESSION BE792513
 VERSION BE792513.1 GI:12097567
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 726)
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 plate: LIM966 row: 1 column: 05
 High quality sequence stop: 724.

FEATURES

SOURCE

Location/Qualifiers

1..726

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4345876"

/tissue_type="adrenal cortex carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_id="NIH_MGC_84"

/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1;
 NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr
 primed. Average insert size 1.229 kb. library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 31.4%; Score 624.6; DB 2; Length 726;
 Best Local Similarity 97.6%; Pred. No. 4.3e-158;

Matches 686; Conservative 0; Mismatches 11; Indels 6; Gaps 5;

QY 867 AAGGCCAAGAGGAGACGCGGTGACGTAACCTTCCCTTACCAAGCTTCCAGCCCCAC 926
 DB 27 AAGGCCAAGAGGAGACGCGGTGAGGTAACTTCCCTTACCAAGCTTCCAGCCCCAC 86
 QY 927 GCCAGCAGCAGCGCTGCTGCCACCCCGTGGCCCGCCAGCAGCTGCTGCCAGGCGAG 986
 DB 87 GCCAGCAGCAGCGCTGCTGCCACCCCGTGGCCCGCCAGCAGCTGCTGCCAGGCGAG 146
 QY 987 AGCCATGCCACATCTGTATATATAGTGGGCTTTTCCAAATACAGCTGCTGTGAAAACT 1046
 DB 147 AGCCATGCCACATCTGTATATATAGTGGGCTTTTCCAAATACAGCTGCTGTGAAAACT 206
 QY 1447 GCATGAACCTCTCTCCGCTGCGCCGCTGCGGGGCTCCAGGCAAGGCCAAGGGGTTG 1106
 DB 207 GCATGAACCTCTCTCCGCTGCGCCGCTGCGGGGCTCCAGGCAAGGCCAAGGGGTTG 266
 QY 1107 GGGGTGGGGCTGTCTCTTCCCTCCACAGGCTGTGTCTTGTGGGGCTGCTCCCATGA 1166
 DB 267 GGTGTGGGGCTGTCTCTTCCCTCCACAGGCTGTGTCTTGTGGGGCTGCTCCCATGA 326
 QY 1167 GACGAGTACCTTAACAGAGTGAAGCCAGGAGATGAGTGGGCTTGGTCTCTGAGG 1226
 DB 327 GACGAGTACCTTAACAGAGTGAAGCCAGGAGATGAGTGGGCTTGGTCTCTGAGG 385
 QY 1227 TTGAGACCCCAAGCTTGGCCACTTCCCTCC-GGCAGTAGCTTCCATCCATCCCTT 1285
 DB 386 TTGAGACCCCAAGCTTGGCCACTTCCCTCCGGGAGTAGCTTCCATCCATCCCTT 445
 QY 1286 CTTTATCTATGATCTATAGGCTCGGTGTGTGTA--ACAACAACACCCCTATGTTTCTC 1343
 DB 446 CTTTATCTATGATCTATAGGCTCGGTGTGTGTAACAACAACACCCCTATGTTTCTC 505

QY 1344 TTCAATATCTACGCAATTACCAATTGGTTGAGGCCAAATTACAGAGCTTCTCAATCAGATT 1403
 DB 506 TTCAATATCTACGCAATTACCAATTGGTTGAGGCCAAATTACAGAGCTTCTCAATCAGATT 565
 QY 1404 TACATCTCCATTTTCATTAACCGGGGAAATCCCGGACCACTGAGTGTGTGTGT 1463
 DB 566 TACATCTCCATTTTCATTAACCGGGGAAATCCCGGACCACTGAGTGTGTGTGT 625
 QY 1464 CACTGAGGTTAGATCTGAACCCAGGGTGTCAACGCTGCTTCACTCCCACTCTGG 1523
 DB 626 CACTGAGGTTAGATCTGAACCCAGGGTGTCAACGCTGCTTCAACCTCCCACTCTGG 683
 QY 1524 GCAGTGGAGATTTTCCCTCATCTTCTTCACTCTCTTAAGGCTAT 1566
 DB 684 GCAGTGGAGATTTTCCCTCATCTTCTTCACTCTCTTAAGGCTAT 726

RESULT 9

CA424992/c

LOCUS

DERIVATION

UI-H-FBI-bee-b-03-0-UI 3', mRNA sequence.

ACCESSION

CA424992

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 672)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: James Martin
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@iowa.edu
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES

source

1..672

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-FBI-bee-b-03-0-UI"

/tissue_type="Cell lines"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/clone_id="NCI CGAP FEL"

/note="Organ: Chondrosarcoma; Vector: pRT73-Pac
 (Pharmacia) with a modified polylinker; Site: 1; EcoR I;
 Site 2: Not I; NCI CGAP_FEL is a normalized cDNA library
 derived from a pool of mRNA obtained from 3 cell lines
 from grade II chondrosarcoma tissues. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pRT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 GCGTACGGAC. The cell lines were provided by Dr James
 Martin from the University of Iowa.
 TAG_TISSUE=human grade 2 chondrosarcoma cell line pool
 TAG_LIB=UI-H-FEL

ORIGIN TAG_SEQ=CGCTACGGAC"

Query Match 30.9%; Score 613.8; DB 6; Length 672;
 Best Local Similarity 98.8%; Pred. No. 3.6e-155;
 Matches 649; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

1271 TCATCATCCCCCTTTATCTATGATCTATAGGCTGGGAGTGTGA--ACAACA 1328
 Db 672 TCATCATCCCCCTTTATCTATGATCTATAGGCTGGGAGTGTGAACAACACA 613

QY 1329 CCCCATCGTGTCTCTCAATACTCAGCATTAACATTTGTTAGGCCAATTAGAGCT 1388
 Db 612 CCCCATCGTGTCTCTCAATACTCAGCATTAACATTTGTTAGGCCAATTAGAGCT 553

QY 1389 TTCTCAATCAGATTATACATCTCCATTTTCATTTAAGGGGAAACATCCCCGAGCCTG 1448
 Db 552 TTCTCAATCAGATTATACATCTCCATTTTCATTTAAGGGGAAACATCCCCGAGCCTG 493

QY 1449 AGTGTGTGTGTGTCTCACTGAAGTTAGATCTGAACCCAGGCTCAACMGCTGTCTCA 1508
 Db 492 AGTGTGTGTGTGTCTCACTGAAGTTAGATCTGAACCCAGGCTCAACMGCTGTCTCA 433

QY 1509 ACTCCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTTAAGGCTATGC 1568
 Db 432 ACTCCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTTAAGGCTATGC 373

QY 1569 ACCCTCTCCCACTCTTCAGCTGGGGGAT--GGGGGAGTCAAGAAAAGCCCCATCT 1627
 Db 372 AACCTCTCCCACTCTTCAGCTGGGGGAT--GGGGGAGTCAAGAAAAGCCCCATCT 313

QY 1628 CCCATCTGGGATAGGAGCTTCATCAGCTTAACCTGGGAAATGCTGCTGCCCAAG 1687
 Db 312 CCCATCTGGGATAGGAGCTTCATCAGCTTAACCTGGGAAATGCTGCTGCCCAAG 253

QY 1688 TGACTCTTGTGTCTCTCCCACTACAGAGAGGGTGAAGGGAGGGTGGCTCAG 1747
 Db 252 TGACTCTTGTGTCTCTCTCCCACTACAGAGAGGGTGAAGGGAGGGTGGCTCAG 193

QY 1748 TTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCCCTCATCTCTGTGCAATGTC 1807
 Db 192 TTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCCCTCATCTCTGTGCAATGTC 133

QY 1808 CTTAGGCTGGCTCTCACTCCCAAGCTCACTGGG--CCCTTGGGGAGGACTGGGAGGGG 1866
 Db 132 CTTAGGCTGGCTCTCACTCCCAAGCTCACTGGG--CCCTTGGGGAGGACTGGGAGGGG 73

QY 1867 CCGTGGAGAGGCTGACGCTGGAACCTGTATACAAATTAAGACAGTCTCAGAGA 1923
 Db 72 CCGTGGAGAGGCTGACGCTGGAACCTGTATACAAATTAAGACAGTCTCAGAGA 16

RESULT 10
 BM996902/c 681 bp mRNA linear EST 26-MAR-2002
 LOCUS UI-H-ED0-awx-c-06-0-UI.s1 NCI CGAP_ED0 Homo sapiens CDNA clone
 DEFINITION IMAGE:5824829 3', mRNA sequence.
 ACCESSION BM996902
 VERSION BM996902.1 GI:19721803
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS Tumor Gene Index
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 COMMENT Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Tissue Procurement: Dr. Jose Mercuende
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>
 Seq primer: M13 Forward
 POLYA=yes

FEATURES
 source Location/Qualifiers
 1..681
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5824829"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP ED0"
 /note="Organ: Left Pubic Bone; Vector: p7773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED0 is a CDNA library containing the following tissue(s): Chondrosarcoma cell line C85. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAAGGCT.
 TAG_TISSUE=chondrosarcoma
 TAG_LIB=UI-H-ED0
 TAG_SEQ=GCTCAAGGCT"

ORIGIN

Query Match 30.9%; Score 613.8; DB 5; Length 681;
 Best Local Similarity 98.8%; Pred. No. 3.7e-155;
 Matches 659; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

1260 GCACTAGCTCTCATCATCATCCCCCTTTATCTATGATCTATAGGCTGGTGTGT 1319
 Db 681 GCACTAGCTCTCATCATCATCCCCCTTTATCTATGATCTATAGGCTGGTGTGT 622

QY 1320 AACA-ACACACCCCTATCGTGTCTTCAATACTCAGATTAACCTTTGGAGGCCAA 1378
 Db 621 AACACACACCCCTATCGTGTCTTCAATACTCAGATTAACCTTTGGAGGCCAA 562

QY 1379 ATTCAGAGCTTTCTCAATCAGATTTTCAATCTCCATTTTCAATTAACGGGAAACATCC 1438
 Db 561 ATTCAGAGCTTTCTCAATCAGATTTTCAATCTCCATTTTCAATTAACGGGAAACATCC 502

QY 1439 CGAGCACTGAGTGTGTGTGTGTCTCACTGAAGTTAGTCTGAACCCAGGCTGCAACN 1498
 Db 501 CGAGCACTGAGTGTGTGTGTGTCTCACTGAAGTTAGTCTGAACCCAGGCTGCAACN 442

QY 1499 GCTGCTCTCAACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTACTCTCT 1558
 Db 441 GCTGCTCTCAACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTACTCTCT 382

QY 1559 AAGGCTATGACCCCTCCCACTCTTCCAGCTGGGGAT--GGGGGAGTCAATAGAAA 1617
 Db 381 AAGGCTATGACCCCTCCCACTCTTCCAGCTGGGGAT--GGGGGAGTCAATAGAAA 322

QY 1618 GGGCCCATCTCCCATCTGGGATAGGAGCTTCATCAGCTTAACCTGGGAAATGCCG 1677
 Db 321 GGGCCCATCTCCCATCTGGGATAGGAGCTTCATCAGCTTAACCTGGGAAATGCCG 262

QY 1678 CTGCCCCAGTGAATCTTGGTTTCGTTCCCATACAGAAACAGAGTGGAGAGGG 1737
 Db 261 CTGCCCCAGTGAATCTTGGTTTCGTTCCCATACAGAAACAGAGTGGAGAGGG 202

QY 1738 TGGGCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTCATGCTCTCTG 1797

Db	201	TGGGCTCAGTTAGAGAGGATCCCCAGGGCAAGTATAGCTCTCCCTCCATAGCCTCTGTG	142
QY	1798	GTCACTGTGCTTAGGTGGTGCCTCTCACTCCACCACTCTTGGG-CCCTTGGGGAGAGACT	1856
Db	141	GTCACTGTGCTTAGGGT-GCTCTCACTCCACCACTCTGGGCCCTTGGGGAGAGACT	83
QY	1857	GGGAGGGGGCCGTGGAGAGCCTCTGACGTGGAACTGTATACCAATAAAGGACAGTC	1916
Db	82	GGGAGGGGGCCGTGGAGAGCCTCTGACGTGGAACTGTATACCAATAAAGGACAGTC	23
QY	1917	TCACAGA 1923	
Db	22	TCACAGA 16	

RESULT	11
BUE624626/c	
LOCUS	BUE624626 676 bp mRNA linear EST 23-SEP-2002
DEFINITION	UI-H-Fg1-bgk-d-20-0-UI s1 NC1 CGAP_Fg1 Homo sapiens CDNA clone
ACCESSION	UI-H-Fg1-bgk-d-20-0-UI 3', mRNA sequence.
VERSION	BUE624626
KEYWORDS	BUE624626.1 GI:23290841
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 676)	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute,	Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg,	Ph.D.		

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strussberg, Ph.D.
Email: cgaabs@omni.nh.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq Primer: W13 FORWARD
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. .676

```

source
1. .676
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-Fgl-Bgk-d-20-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="MDH03 (Life Technologies)"
/clone_lib="NCI_CGAP_Fgl"
/note="Organ: Enchondroma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Fgl is a normalized cDNA library obtained from a pool of mRNA from 2 cell lines from Enchondroma tissues. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First stranded cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGGTCACTC. The cell lines were provided by Dr. James Martin from the University of Iowa. TAG-TISUF-Enchondroma cell line (Mix of EN1 and EN2) TAG-Lib-UI-H-Fgl TAG_SEQ=CGGTCACTC"

```

ORIGIN	Query Match
--------	-------------

30.9%; Score 613.2; DB 5; Length 676;

Best Local Similarity 99.1%; Pred. No. 5,3e-155;
Matches 637; Conservative 0; Mismatches 4; Indels 2; Gaps 2

OY	1283	CCCTTTAATCTAAGATCTAATAGAGCTCGGAGTGTGAACAACACAC-CCCTATCGTGT	1341
Db	660	CCCTTTAATCTAAGATCTAATAGAGCTCGGAGTGTGAACAACAACACCCCTATCGTGT	601
OY	1342	CCCTCAAAATCTACGCAATTACCAATTGGTTGAGGCGCAATTCAAGCTTTCTCAATACGA	1401
Db	600	CCCTCAAAATCTACGCAATTACCAATTGGTTGAGGCGCAAAATTCAGAGCTTTCTCAAAATACGA	541
OY	1402	TTTCAATATCTCCATTTTCATTTAACGGGGAAAACATCCCCGAGCCATGAGTCTGTGCTT	1461
Db	540	TTTCAATATCTCCATTTTCATTTAACGGGGAAAACATCCCCGAGCCATGAGTCTGTGCTT	481
OY	1462	GTCATGAAGGTTGATCTGAAGCCAGAGGTGTCAAGAGCTGCTCAACTGCCACCTCT	1521
Db	480	GTCATGAAGGTTGATCTGAAGCCAGAGGTGTCAACAGCTGCTCTCAACTGCCACCTCT	421
OY	1522	GGGCACTGAGAGATTTTCCCTCATTTCTACCTCTAAGGCTATGCAACCCCTCCACG	1581
Db	420	GGGCACTGAGAGATTTTCCCTCATTTCTACCTCTAAGGCTATGCAACCCCTCCACG	361
OY	1582	TCTTCACAGCTGGGGGATGGGGGAGGTATAGGAAAACCCCACTCCCATCTGGAGATG	1641
Db	360	TCTTCACAGCTGGGGGATGGGGGAGGTATAGGAAAACCCCACTCCCATCTGGAGATG	301
OY	1642	GGACCTTCOCATCAGCCTTTAACCCCTGGGAAATGCTGCTGCCCCCAAGTACTTTGGTTTC	1701
Db	300	GGACCTTCOCATCAGCCTTTAACCCCTGGGAAATGCTGCTGCCCCCAAGTACTTTGGTTTC	241
OY	1702	GTCCTCCACATATCAGAAGCAGAGGTGAGAGGGGAAAGGAGTCTCAGTTAGCAGAGGGTCC	1761
Db	240	GTCCTCCACATATCAGAAGCAGAGGTGAGAGGGGAAAGGAGTCTCAGTTAGCAGAGGGTCC	181
OY	1762	CAGGGCAAGTCAGGCTCTCTCCCTCAGTGCCTCTTGGTCAATGTGCTTAGAGGTGCTTC	1821
Db	180	CAGGGCAAGTCAGGCTCTCTCCCTCAGTGCCTCTTGGTCAATGTGCTTAGAGGTGCTTC	121
OY	1822	TCACTCCACCACTCTCTGGG-CCCTTGGGAGGAGCTGGGAGGAGGAGCGTGGAGAGCC	1880
Db	120	TCACTCCACCACTCTCTGGGCGCCCTTGGGAGGAGAGCTGGGAGGAGGAGCGTGGAGAGCC	61
OY	1881	TGAGCGTGAACCTGTATACACATAAAGCAGAGCTTCACAGA	1923
Db	60	TGAGCGTGAACCTGTATACACATAAAGCAGAGCTTCACAGA	18

RESULT 12
BF306561
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF306561 636 bp mRNA linear EST 21-NOV-2000
601888988F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122758 5',
mRNA sequence.
BF306561
BF306561.1 GI:11253663
EST.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 636)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINT at: image.llnl.gov
Plate: LHCMT006 row: 1 column: 21

QY 431 TCCTAAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGAGCCGCCAGC 490
Db 193 TCCTAAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGAGCCGCCAGC 252
QY 491 CCCAGACCCCAAGGTGACATCTCTCCCAACTAGAGGCTCTCACTCTATCCAGGCAAGC 550
Db 253 CCCAGACCCCAAGGTGACATCTCTCCCAACTAGAGGCTCTCACTCTATCCAGGCAAGC 312
QY 551 CAGGAGCTCCCTGCTGACACATGATGCCAGATTTCAGATTGGCTCCGTCACCTTA 610
Db 313 CAGGAGCTCCCTGCTGACACATGATGCCAGATTTCAGATTGGCTCCGTCACCTTA 372
QY 611 TCAGAGTACAGGGGCTGAGGCTCAGGAGAGAGATCTTAAGAACCCACTGTGGTCAAG 670
Db 373 TCAGAGTACAGGGGCTGAGGCTCAGGAGAGAGATCTTAAGAACCCACTGTGGTCAAG 432
QY 671 GGAATGGAGACCAAGAGACATATGGGCAAGCTCTGACAGACAGACAGACAAACCTT 730
Db 433 GGAATGGAGACCAAGAGACATATGGGCAAGCTCTGACAGACAGACAGACAAACCTT 492
QY 731 CTGATCTATGAGTCTCTGACAGGCAAGGAGACCAAGCTGAACTCTTGGCCAG 790
Db 493 CTGATCTATGAGTCTCTGACAGGCAAGGAGACCAAGCTGAACTCTTGGCCAG 552
QY 791 GGGAGTGGAGAGACAGAGGAAAGTCAACAGGAGGCTGCTATCTTAAGTGAATTAAT 850
Db 553 GGGAGTGGAGAGACAGAGGAAAGTCAACAGGAGGCTGCTATCTTAAGTGAATTAAT 612
QY 851 TGCCCGAGGCTCAGACAGGCCAAGAGAGACAGCCCTGACG 892
Db 613 TGCCCGAGGCTCAGACAGGCCAAGAGAGACAGCCCTGACG 654

RESULT 14
BF203719 976 bp mRNA linear EST 06-NOV-2000
LOCUS 601865380P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:409458 5',
DEFINITION mRNA sequence.
ACCESSION BF203719
VERSION BF203719.1 GI:11097305
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 976)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gsa@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LINC967 row: P column: 19
High quality sequence stop: 676.
Location/Qualifiers
1. 976
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:409458"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7; Site: 1; EcotRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by

ORIGIN
Query Match 29.6%; Score 587.4; DB 2; Length 976;
Best Local Similarity 93.1%; Pred. No. 6e-148;
Matches 712; Conservative 0; Mismatches 43; Indels 10; Gaps 9;
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

QY 761 GACCAAGGACCTGAAACCTCTTGCCCAAGGGAGTGGGAGAGACAGAGGGAGGTCA 820
Db 1 GACCAAGGACCTGAAACCTCTTGCCCAAGGGAGTGGGAGAGACAGAGGGAGGTCA 59
QY 821 GGCAGAGGTGCTTATCTAAGTGAATTAATGCTCCGAGGCTCAGCAAGCCAGAGAG 880
Db 60 GGCAGAGGTGCTTATCTAAGTGAATTAATGCTCCGAGGCTCAGCAAGCCAGAGAG 119
QY 881 ACAGCCGTGACGGTAACTTCCCTCTACAGACCTCCAAAGCCCAAGGAGAGAGC 940
Db 120 ACAGCCGTGACGGTAACTTCCCTCTACAGACCTCCAAAGCCCAAGGAGAGAGC 179
QY 941 TGCCTGCCACCCCGTGCACAGCCAGCTGCTGACAGGAGAGACCATGACATC 1000
Db 180 TGCCTGCCACCCCGTGCACAGCCAGCTGCTGACAGGAGAGACCATGACATC 238
QY 1001 TGTATATGATGGGCTTTTCCATATACAGCTGCTGCTGAAATACTGATGAATCTCTG 1060
Db 239 TGTATATGATGGGCTTTTCCATATACAGCTGCTGCTGAAATACTGATGAATCTCTG 298
QY 1061 CGTCTGCGCTGCTGCGGGGCTCCAGGAGAGGCCAAGGGGCTGGGGGCTGGT 1120
Db 299 CGTCTGCGCTGCTGCGGGGCTCCAGGAGAGGCCAAGGGGCTGGGGGCTGGT 358
QY 1121 CCTTCTCCCTCCACAGGCTGTGTTCTNTGGGCTGCTCCATGAGACAGATCACCTA 1180
Db 359 CCTTCTCCCTCCACAGGCTGTGTTCTNTGGGCTGCTCCATGAGACAGATCACCTA 418
QY 1181 ACAGAGATGAGACCAAGGCAATGATGGGCTTGGGCTCTCGAGCTGGACCCAGCTT 1240
Db 419 ACAGAGATGAGACCAAGGCAATGATGGGCTTGGGCTCTCGAGCTGGACCCAGCTT 476
QY 1241 CTGACACCTTCCCTCCAGGCTGTGTTCTNTGGGCTGCTCCATGAGACAGATCACCTA 1299
Db 477 CTGACACCTTCCCTCCAGGCTGTGTTCTNTGGGCTGCTCCATGAGACAGATCACCTA 536
QY 1300 TCTATATGCTCGGTGTGTATACCAACACACCTATGCTTGTCTTCAATATCTAGAT 1359
Db 537 TCTATATGCTCGGTGTGTATACCAACACACCTATGCTTGTCTTCAATATCTAGAT 595
QY 1360 TACCATGCTTGGAGCCAAATCAGAGCTTCTCAATACAGATTACATCTCCATTTTC 1419
Db 596 TACCATGCTTGGAGCCAAATCAGAGCTTCTCAATACAGATTACATCTCCATTTTC 653
QY 1420 ATTAACTGGGAAACATCCCGAGACCACTGAGTGTCTGTCTTGTATCACTGAAGTTATC 1479
Db 654 ATTAACTGGGAAACATCCCGAGACCACTGAGTGTCTGTCTTGTATCACTGAAGTTATC 711
QY 1480 TGAACCCAGGGGTCAACAGCTGCTCAAGCTCCCACTCTGGG 1524
Db 712 TGAACCCAGGGGTCAACAGCTGCTCAAGCTCCCACTCTGGG 756

RESULT 15
BF203383 956 bp mRNA linear EST 06-NOV-2000
LOCUS 601865395P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:409646 5',
DEFINITION mRNA sequence.
ACCESSION BF203383
VERSION BF203383.1 GI:11096969
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 956)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gsa@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LINC967 row: P column: 19
High quality sequence stop: 956.
Location/Qualifiers
1. 956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:409646"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7; Site: 1; EcotRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by

REFERENCE 1 (bases 1 to 956)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: image.lnl.gov
Plate: LLCM965 row: n column: 23
High quality sequence stop: 682.
Location/Qualifiers

FEATURES

Source

1..956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMGR:409846"
/tissue_type="Thadomysarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI; Site: 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 29.4%; Score 585; DB 2; Length 956;
Best Local Similarity 89.5%; Pred. No. 2.7e-147;
Matches 708; Conservative 0; Mismatches 72; Indels 11; Gaps 7;
QY 766 GGGACCTGGAGACCTTTGGCCAGGAGTGGAGAGACAGAGGAGTACAGGCA 825
DB 1 GGGACCTGGAGACCTTTGGCCAGGAGTGGAGAGACAGAGGAGTACAGGCA 60
QY 826 GGGTGGCTTCTTAAGTGAACCTAATTCGCCAGGCGTCAAGGCCAAGAGACAGC 885
DB 61 GGGTGGCTTCTTAAGTGAACCTAATTCGCCAGGCGTCAAGGCCAAGAGACAGC 120
QY 886 CGTAGAGGTAACCTCCCTTACAGAGCTCCAGGCCAGCGAGCGAGGCTGCT 945
DB 121 CGTAGAGGTAACCTCCCTTACAGAGCTCCAGGCCAGCGAGCGAGGCTGCT 180
QY 946 GCCCAGCCCGTGCCTCCAGCCAGCTGCTGTCAGAGGCAAGCCATCTGTAT 1005
DB 181 GCCCAGCCCGTGCCTCCAGCCAGCTGCTGTCAGAGCCATCTGTAT 239
QY 1006 ATAGATGGGGTTTTTCCAAATACAGCTGGTGTGAAAACTGATGAATCTGCCGTC 1065
DB 240 ATAGATGGGGTTTTTCCAAATACAGCTGGTGTGAAAACTGATGAATCTGCCGTC 299
QY 1066 CTGCGCTGCTGGGGCTTCAGGCAAGGCAAGTGGGGTGGGGCTGTGCTCTTC 1125
DB 300 CTGCGCTGCTGGGGCTTCAGGCAAGGCAAGTGGGGTGGGGCTGTGCTCTTC 359
QY 1126 TCCCTCCCAAGGCTGTGTCTCTGGGGTGTCTCCATGCAAGATCACTTAACGA 1185
DB 360 TCCCTCCCAAGGCTGTGTCTCTGGGGTGTCTCCATGCAAGATCACTTAACGA 419
QY 1186 GATGAGAGCAGGCGATGATGGGCTTTGGTCTCTGAGGTGAGCCCAAGCTTCTGC 1245
DB 420 GATGAGAGCA-GGATGATGGGCTTTGGTCTCTGAGGTGAGCCCAAGCTTCTGC 478
QY 1246 CACTTCCCTCCGCGAGTCACTTCATCCATCCCTCTTAACTTAATGAATCTATA 1305
DB 479 CACTTCCCTCCGCGAGTCACTTCATCCATCCCTCTTAACTTAATGAATCTATA 538

QY 1306 GGTCTGGTGTGTAAACAACACACCCCTATGTTGCTTCAAAATCTAGCATTAACAT 1365
DB 539 GGCTGGGTGTGTAAACAACACA-CCCTATGTTGCTTCAAAATCTAG-ATTAACT 596
QY 1366 TGGTGAAGCCAAATTCAGAGCTTTCCTCAATGATTAATCAATCTCCATTTCAATAC 1425
DB 597 TGGTGAAGCCAAATTCAGAGCTTTCCTCAATGATTAATCAATCTCCATTTCAATAC 656
QY 1426 GGGGAAACATCCCGAGCCACTGAGTGTGCTTTGTCACTGAAGATTAGATCTGAAC 1485
DB 657 GG-----ACATCCGAGCACTAGTG-TGTGCTATGTCAAGTGAAGATTAGATCTG-ACC 709
QY 1486 CAGGTGTCAAGTGTCTCTCACTCCCACTCGGGCACTGAGAGATTTCCCTC 1545
DB 710 CAGGTGTCAAGTGTCTTAACTCCCACTCGGGCACTGAGAGATTTCCCTTATCTAC 769
QY 1546 ATTCTACCTCT 1556
DB 770 TTTTATGTTT 780

Search completed: October 24, 2004, 15:41:05
Job time : 6400 secs

QY 121 CCGAAGACATCATTTGTCAACACCTCCAAACAAACCAATCGAGGTGGGGGTGAGAA 180
Db 121 CCGAAGACATCATTTGTCAACACCTCCAAACAAACCAATCGAGGTGGGGGTGAGAA 180
QY 181 AGGTGCGGTGTAACGGAACATGTATCATGAAAACACCTTACGACTACAAAGTGGCAACT 240
Db 181 AGGTGCGGTGTAACGGAACATGTATCATGAAAACACCTTACGACTACAAAGTGGCAACT 240
QY 241 GCGGAGAGAGTGGAGACCCGAGAGTGGGTGACCTTGGCTACTGTGAGAGAGACGAGCAG 300
Db 241 GCGGAGAGAGTGGAGACCCGAGAGTGGGTGACCTTGGCTACTGTGAGAGAGAGCAGCAG 300
QY 301 CCTCACTAGTCCGGGACGCGTACCCGATACAGAACAGTCCAGACAGACTCTCCGAC 360
Db 301 CCTCACTAGTCCGGGACGCGTACCCGATACAGAACAGTCCAGACAGACTCTCCGAC 360
QY 361 GGAAGATCAAAATTTGAGTGGCTCTCTCCCTTCCCTTCCCTGTCGCCGCCGCCAGCTGC 420
Db 361 GGAAGATCAAAATTTGAGTGGCTCTCTCCCTTCCCTTCCCTGTCGCCGCCGCCAGCTGC 420
QY 421 CAGCAAGGCTGCTTAACCCCATTTACAGAGCTCCAGAGCATCTGAGCCAGGTTCTAGC 480
Db 421 CAGCAAGGCTGCTTAACCCCATTTACAGAGCTCCAGAGCATCTGAGCCAGGTTCTAGC 480
QY 481 CCCACAGCACCCCGAGACCCGAGGTGACCATCTTCCAAACTAGGGCCCTCCACTATTC 540
Db 481 CCCACAGCACCCCGAGACCCGAGGTGACCATCTTCCAAACTAGGGCCCTCCACTATTC 540
QY 541 CAGGGAGGCGAGGGAGCTCCCTGAGCTGACATATGATGCCAGATTTGAGTTGGGCTC 600
Db 541 CAGGGAGGCGAGGGAGCTCCCTGAGCTGACATATGATGCCAGATTTGAGTTGGGCTC 600
QY 601 CCTCACTTAATCAGAGATACAGAGGCTGGGATCAGGAGAGAGAAATTTAAAGAACCACT 660
Db 601 CCTCACTTAATCAGAGATACAGAGGCTGGGATCAGGAGAGAGAAATTTAAAGAACCACT 660
QY 661 GTGGGTGAGGGGAAATGGGACAGACATATGGGCAAGCTCTGACAGACAGACAGACA 720
Db 661 GTGGGTGAGGGGAAATGGGACAGACATATGGGCAAGCTCTGACAGACAGACAGACA 720
QY 721 GACAAACCTCTGATCTATGAAAGTCTGACAGGGGCAAGGGGACAGGAGCTGGAACTCT 780
Db 721 GACAAACCTCTGATCTATGAAAGTCTGACAGGGGCAAGGGGACAGGAGCTGGAACTCT 780
QY 781 CTGGGCAAGGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGGTGTCTATTAAG 840
Db 781 CTGGGCAAGGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGGTGTCTATTAAG 840
QY 841 TGAACCTAATTTGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGTAACCTT 900
Db 841 TGAACCTAATTTGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGTAACCTT 900
QY 901 CCCCTCTACAGCTCTCAAGGCCCAAGCCAGCAGAGCAGGCTGCTGCCACCCCGTGC 960
Db 901 CCCCTCTACAGCTCTCAAGGCCCAAGCCAGCAGAGCAGGCTGCTGCCACCCCGTGC 960
QY 961 CCAGCAGCTGCTGTGCGAGGGGAGAGCCATGCCATCTGTANTATGATGGGTTTTT 1020
Db 961 CCAGCAGCTGCTGTGCGAGGGGAGAGCCATGCCATCTGTANTATGATGGGTTTTT 1020
QY 1021 CCATATACAGCTGTGTCTGTAAAAACTGATGAACTCTGCGCTGCTGCGCTGCTG 1080
Db 1021 CCATATACAGCTGTGTCTGTAAAAACTGATGAACTCTGCGCTGCTGCGCTGCTG 1080
QY 1081 CCTCAAGGCAAGGCGCAAGTGGGGTGGGGGTGGGGGTGGGTCTCTTCTCCCAAGGCC 1140
Db 1081 CCTCAAGGCAAGGCGCAAGTGGGGTGGGGGTGGGGGTGGGTCTCTTCTCCCAAGGCC 1140
QY 1141 TGTGTTCTGAGGGCTGCTCCCATGAGACAGATCACTTAACAGAGTGAAGCCAGGCC 1200
Db 1141 TGTGTTCTGAGGGCTGCTCCCATGAGACAGATCACTTAACAGAGTGAAGCCAGGCC 1200

QY 1201 ATGATGGGGCTTTGGGTCTCGAGGTGGAGCCCAAGCTTCTTGGCACTTCCCTCCGG 1260
Db 1201 ATGATGGGGCTTTGGGTCTCGAGGTGGAGCCCAAGCTTCTTGGCACTTCCCTCCGG 1260
QY 1261 CAGTACGCTCTCCATCCATCCCTCCCTCTTATATGAAATCTATAGGCTCGGTGTGTA 1320
Db 1261 CAGTACGCTCTCCATCCATCCCTCCCTCTTATATGAAATCTATAGGCTCGGTGTGTA 1320
QY 1321 ACAACACACCCCTATGTGTGTCTTCAAAATCTACAGATTTACATTTGGTTAGGCCAAAT 1380
Db 1321 ACAACACACCCCTATGTGTGTGTCTTCAAAATCTACAGATTTACATTTGGTTAGGCCAAAT 1380
QY 1381 TCAGAGCTTCTCAAAATCAAGATTTACAAATCTCCATTTGATTAACGGGAAAACATCCCG 1440
Db 1381 TCAGAGCTTCTCAAAATCAAGATTTACAAATCTCCATTTGATTAACGGGAAAACATCCCG 1440
QY 1441 AGCCACTGAGTGTGTGTCTTGTCTTCACTGAAAGTTAATCTGAAATCCAGGGTGTCAAACGC 1500
Db 1441 AGCCACTGAGTGTGTGTGTCTTGTCTTCACTGAAAGTTAATCTGAAATCCAGGGTGTCAAACGC 1500
QY 1501 TGCTCTCAACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTACTCTCTAA 1560
Db 1501 TGCTCTCAACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTACTCTCTAA 1560
QY 1561 GGCATATGACCTCTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTACTCTCTAA 1620
Db 1561 GGCATATGACCTCTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTACTCTCTAA 1620
QY 1621 CCCATCTCCATTTGGGATAGGGAATCTTCACTGAGCTTAACTTGAAGGAGGATGCTGCTG 1680
Db 1621 CCCATCTCCATTTGGGATAGGGAATCTTCACTGAGCTTAACTTGAAGGAGGATGCTGCTG 1680
QY 1681 CCCCAGTGACTCTTGTGTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 1740
Db 1681 CCCCAGTGACTCTTGTGTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 1740
QY 1741 GTCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTTCACTGCTCTGTGTC 1800
Db 1741 GTCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTTCACTGCTCTGTGTC 1800
QY 1801 AGTGTCTCTTAAAGGCTGCTCTCACTCCCACTCTGAGGCTTGGGGGAGAGTGGGG 1860
Db 1801 AGTGTCTCTTAAAGGCTGCTCTCACTCCCACTCTGAGGCTTGGGGGAGAGTGGGG 1860
QY 1861 AGGGGCTCTGGGAGAGGCTCTGACCTGGAACCTGTATACAAATTAAGAGACAGCTCAC 1920
Db 1861 AGGGGCTCTGGGAGAGGCTCTGACCTGGAACCTGTATACAAATTAAGAGACAGCTCAC 1920
QY 1921 AGACNTCTGAGAGGCTCTGACCTGACCTGCACTTCAAACTTGGGAGAGGCTTACTTG 1980
Db 1921 AGACNTCTGAGAGGCTCTGACCTGACCTGCACTTCAAACTTGGGAGAGGCTTACTTG 1980
QY 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 2
AX017502 2192 bp DNA linear PAT 07-SEP-2000
LOCUS AX017502
DEFINITION Sequence 48 from Patent WO9947655.
ACCESSION AX017502
VERSION AX017502.1 GI:10042299
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Schmitt,A., Specht,T., Dahl,B., Hinzmann,B., Rosenthal,A. and Pilsarsky,C.
Human nucleic acid sequences from normal breast tissue
Patent: WO 9947655-A 48 23-SEP-1999;
JOURNAL

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
Location/Qualifiers
1. 2192
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 84.8%; Score 1684.4; DB 6; Length 2192;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1859; Conservative 0; Mismatches 73; Indels 24; Gaps 11;

33 GCCCATCCCAACCAAGCCCGGTGGGCGAGCAATCATMAAGCCCTAGAGAGCGCTATGA 92
258 GGCCTTCCAGCCCGCCCGGTGGGCGAGCAATCATMAAGCCCTAGAGAGCGCTATGA 317
93 GTTTGGGGTGGAGGTGAGACTTCTCACTGAAAGCATCATGTGACCAACCTCCAAAGA 152
318 GTTTGGGGTGGAGGTGAGACTTCTCACTGAAAGCATCATGTGACCAACCTCCAAAGA 374
153 AACCAATCATGAGGTGGGGCTGAGAAAGCTGGCGGTGAACGGAACCTGTACATGAAA 212
375 CAACCAATCATGAGGTGGGGCTGAGAAAGCTGGCGGTGAACGGAACCTGTACATGAAA 432
213 CAACCAATCATGAGGTGGGGCTGAGAAAGCTGGCGGTGAACGGAACCTGTACATGAAA 272
433 TTGGCT-----CACAAAGTCCAG-CTGGGAGAGAGCTGAGCC--GACGTGGTGA 481
273 CCTGGCTACTGTGCGAGAGAGCGAGAGCTCTCACTGTCGGGCGAGCGTCAACCCGATA 332
482 CTTGG--CTGGCGGAGAGAGCGAGAGCTCTCACTGTCGGGCGAGCGTCAACCCGATA 538
333 CAGAACAGTTCAGAGAGAGCTTTCGGAAGGAGATCAAAATCTAGTGGCTCTCTCTCC 392
539 CAGAACAGTTCAGAGAGAGCTTTCGGAAGGAGATCAAAATCTAGTGGCTCTCTCTCC 598
393 TTTTCCCTGT-CCCCCGCCCGAGCGCTGAGCAAAAGCTCGTAAACCCATTACAAG 451
599 TTTTCCCTGT-CCCCCGCCCGAGCGCTGAGCAAAAGCTCGTAAACCCATTACAAG 658
452 CTCAGAGACATCTAGCCAGGTTTAAAGCCCAAGCAAGCCCAAGCCAGGTGAGCAAT 511
659 CTCAGAGACATCTAGCCAGGTTTAAAGCCCAAGCAAGCCCAAGCCAGGTGAGCAAT 718
512 CTCCTCAACTAGAGGCGCTCACTATTCAGAGGAGCGGCGAGGAGCTCCCTGGCTGAGA 571
719 CTCCTCAACTAGAGGCGCTCACTATTCAGAGGAGCGGCGAGGAGCTCCCTGGCTGAGA 778
572 CATGATGCCAGATTTTCAATTTGGCTCCGTCATTTAATTCAGAGTACAGGGGCTGGG 631
779 CATGATGCCAGATTTTCAATTTGGCTCCGTCATTTAATTCAGAGTACAGGGGCTGGG 838
632 TCAGAGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGGAGTGGGACAGAGACAT 691
839 TCAGAGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGGAGTGGGACAGAGACAT 898
692 ATGGGCAAGCTCTGCAAG 751
899 ATGGGCAAGCTCTGCAAG 958
752 GGGCAAGGGAG 811
959 GGGCAAGGGAG 1018
812 AAGGTCAAG 871
1019 AAGGTCAAG 1078
872 CAAG 931
1079 CAAG 1138

922 CGAGCAGGCTCCTGCGCCACCCCGTGGCCGAGCAGTGGCTGTGCGAGGCGAGAGCCA 991
1139 CGAGCAGGCTCCTGCGCCACCCCGTGGCCGAGCAGTGGCTGTGCGAGGCGAGAGCCA 1198
992 TGCCCATCTGTAT 1051
1199 TGCCCATCTGTAT 1258
1052 AAATCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 1111
1259 AAATCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 1318
1112 GGGGTGGTCTTCT 1171
1319 GGGGTGGTCTTCT 1378
1172 GATCACTAAGAGAGATGAG 1231
1379 GATCACTAAGAGAGATGAG 1438
1232 CCCCAGCTTCTTCCCACTTCCCTCC-GGCAGTCACTCTCCATCCATCCCTCTCTTA 1290
1439 CCCCAGCTTCTTCCCACTTCCCTCC-GGCAGTCACTCTCCATCCATCCCTCTCTTA 1498
1291 ATCTATGATCTATATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1349
1499 ATCTATGATCTATATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1558
1350 TACTAGCATTAACATTTGTTGAGGCGCAATTCAGAGCTTTCTCAATTCAGATTTACAT 1409
1559 TACTAGCATTAACATTTGTTGAGGCGCAATTCAGAGCTTTCTCAATTCAGATTTACAT 1618
1410 CTCGATTTTCAATTAAGGGGAGAAACATCCCGAGGACAGAGAGAGAGAGAGAGAGAG 1469
1619 CTCGATTTTCAATTAAGGGGAGAAACATCCCGAGGACAGAGAGAGAGAGAGAGAGAG 1678
1470 AGGTATGATCTGAAACCCAGGAGTGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1529
1679 AGGTATGATCTGAAACCCAGGAGTGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1738
1530 AGGATATTTTCCCTCATTTTCACTCTTAAAGGCTATGCAACCTCCCGAGGCTTCCAG 1589
1739 AGGATATTTTCCCTCATTTTCACTCTTAAAGGCTATGCAACCTCCCGAGGCTTCCAG 1798
1590 CTGGGGAGATGGGGAGATCATAGAAAGCCCGCATCTCCATCTGGAGTGGGAGCTTC 1649
1799 CTGGGGAGATGGGGAGATCATAGAAAGCCCGCATCTCCATCTGGAGTGGGAGCTTC 1858
1650 CATGACCTTAAACCTGCGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1709
1859 CATGACCTTAAACCTGCGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1918
1710 CATAG 1769
1919 CATAG 1978
1770 GTCAAGCTTCTCTCTCAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1829
1979 GTCAAGCTTCTCTCTCAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2038
1830 AACACTCTGGG-CCCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
2039 AACACTCTGGG-CCCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
1889 GAACCTGTATACATATTAAG 1948
2099 GAACCTGTATACATATTAAG 2157
1949 TCTCAAACTTAAAG 1983
2158 TCTCAAACTTAAAG 2192

RESULT 3
BD135192 2192 bp DNA linear PAT 18-SEP-2002
LOCUS Human nucleic acid sequence originating in normal mammary tissue.
DEFINITION
ACCESSION BD135192
VERSION BD135192.1 GI:23230137
KEYWORDS JP 2002506639-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2192)
Speft,T., Hintzman,B., Shcmitt,A., Pirarski,C., Duhl,E. and
Rosenthal,A.
Human nucleic acid sequence originating in normal mammary tissue
Patient: JP 2002506639-A 39 05-MAR-2002;
JOURNAL METAGEN GESELTSCHAFT FUER GENOME FORSCHUNG MBH
COMMENT
CS Homo sapiens (human)
PN JP 2002506639-A/39
PD 05-MAR-2002
PF 19-MAR-1999 JP 2000536838
PI 20-MAR-1998 DE 198 13 835 0
PI THOMAS SPEFT, BERND HINTZMAN, ARMIN SCHMITT, CHRISTIAN PIRARSKI,
PI EDGAR DUHL,
PI ANDRE ROSENTHAL
PC C12N15/09,A61K48/00,A61P35/00,A61P43/00,A61P43/00,C07K14/47,
PC C07K16/18,
PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68//A61K38/00, PC
C12N15/00,
PC C12N5/00,A61K37/02
CC Human nucleic acid sequence originating in normal mammary tissue
FH Key Location/Qualifiers
FT source 1. 2192
FT /organism='Homo sapiens (human)'.
1. 2192
Location/Qualifiers
/organism='Homo sapiens'
/mol_type='Genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 84.8%; Score 1684.4; DB 6; Length 2192;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1858; Conservative 0; Mismatches 73; Indels 24; Gaps 11;
QY 33 GCCCATCCCAACCAAGCCCGGTGGGCAAGCAATCAAGACCTTAGAGAGCGCTATGA 92
DB 258 GGCCTTCCCAAGCCCGCGGTGGGCAAGCAATCAAGACCTTAGAGAGCGCTATGA 317
QY 93 GTTGGGCGTGAAGTGAAGACTTCTCACTGAAGATCAATGTCACCACTCCCAACA 152
DB 318 GTTGGGCGTGAAGTGAAGACTTCTCACTGAAGATCAATGTCACCACTCCCAACA 374
QY 153 AACCAACATGAGTGCAGGAGCTGAGAAAGCTGGGCGTGAACGAGACCTGTACATGAAA 212
DB 375 CAACCAATGAGTGCAGGAGCTGAGAAAGCTGGGCGTGAACGAGACCTGTACATGAAA 432
QY 213 CACCTTAGACATCAACAATGCGGAGCTGCGGAGAGAGTGAACCCGAGCGTGGGTGA 272
DB 433 TTGGCT-----CAAGATGCCAG-CTGCGGAGAGAGTGAACCC--GACGTGCGTGA 481
QY 273 CCTGGGCTAGTGCAGGAGAGAGAGAGAGTCACTAGTCCGGGAGCGGCTCAACCGCATGA 332
DB 482 CCGG-----CTTGGGAGAGAGAGAGAGAGTCACTAGTCCGGGAGCGGCTCAACCGCATGA 538
QY 333 CAGAAACGCTCCAGAGAGCTTCCGAGCGAGAGATCAAAATCTAGTGCCTTCCCTTCCC 392
DB 539 CAGAAACGCTCCAGAGAGCTTCCGAGCGAGAGATCAAAATCTAGTGCCTTCCCTTCCC 598
QY 393 TTTCCTGT-CCCCCGGCGGAGCGCTGCGAGCAAGCTGCTTAACCCATTCAACAG 451
|||||

DB 599 TTTCCTGTGCCCCCGGCGGAGCGCTGCGAGCAAGAGCTGCGTAACCCATTCAACAG 658
QY 452 CTCAGAGACATCTCAGAGAGCTTTAGCCCCCAGCAGCCAGAGAGAGAGAGAGAG 511
DB 659 CTCAGAGACATCTCAGAGAGCTTTAGCCCCCAGCAGCCAGAGAGAGAGAGAGAG 718
QY 512 CTTCCAAACTAGAGGAGCTTCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 571
DB 719 CTTCCAAACTAGAGGAGCTTCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 778
QY 572 CATGATGCCCAATTTCAATTTGGGCTCCGTACTTAATCCAGAGATACAGAGGAGGAG 631
DB 779 CATGATGCCCAATTTCAATTTGGGCTCCGTACTTAATCCAGAGATACAGAGGAGGAG 838
QY 632 TCAGAGAGAGAAATCTTAAGAACCACTGTGGGTGAGGAGGAGGAGGAGGAGGAGGAG 691
DB 839 TCAGAGAGAGAAATCTTAAGAACCACTGTGGGTGAGGAGGAGGAGGAGGAGGAGGAG 898
QY 692 ATGGGCAAGCTCTGACAGAGACAGACAGACAGACAAACCTCTGTATCTAAGTCTTGCA 751
DB 899 ATGGGCAAGCTCTGACAGAGACAGACAGACAGACAAACCTCTGTATCTAAGTCTTGCA 958
QY 752 GGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 811
DB 959 GGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1018
QY 812 AAGGTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 871
DB 1019 AAGGTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1078
QY 872 CAAGAGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 931
DB 1079 CAAGAGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138
QY 932 CGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991
DB 1139 CGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1198
QY 992 TGGCAATCTGATATAGATGGGAGTTTTCATATAGAGTGTGTGAAAAATGCAAG 1051
DB 1199 TGGCAATCTGATATAGATGGGAGTTTTCATATAGAGTGTGTGAAAAATGCAAG 1258
QY 1052 AAATCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGG 1111
DB 1259 AAATCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGG 1318
QY 1112 GGGGCTGCTCTTCT 1171
DB 1319 GGGGCTGCTCTTCT 1378
QY 1172 GATCACTTAACAGAGATGGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1231
DB 1379 GATCACTTAACAGAGATGGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1438
QY 1232 CCGCAGCTTCTGCACTTCCCTCTCC-GGCAGTCACTTCCATCATCCCTCTCTTTTA 1290
DB 1439 CCGCAGCTTCTGCACTTCCCTCTCCGAGGAGTCACTTCCATCATCCCTCTCTTTTA 1498
QY 1291 ATCTAGATCTATAGGCTCGGAGTGTGAAGA-ACAGACCCCTAATCGTGTCTTCAAA 1349
DB 1499 ATCTAGATCTATAGGCTCGGAGTGTGAAGA-ACAGACCCCTAATCGTGTCTTCAAA 1558
QY 1350 TACTCAGATTAACATTTGTTGAGGAGCAATTCAGAGCTTTCTCAATCAGATTAAAT 1409
DB 1559 TACTCAGATTAACATTTGTTGAGGAGCAATTCAGAGCTTTCTCAATCAGATTAAAT 1618
QY 1410 CTCATTTTCTATTAAAGGGGAAAACATCCCGGAGCCATGAGTGTGTCTTCTCACTGA 1469
DB 1619 CTCATTTTCTATTAAAGGGGAAAACATCCCGGAGCCATGAGTGTGTCTTCTCACTGA 1678
QY 1470 AGGTTAGATCTGAACCAAGGAGTCAACAGTGTCTCAACTCCCACTCTGGGAGACTG 1529
DB 1679 AGGTTAGATCTGAACCAAGGAGTCAACAGTGTCTCAACTCCCACTCTGGGAGACTG 1738

QY 1530 AGGAGTATTTTCCCTCTTCTTACCTCTCTAAGGCTATGACCCCTCCACGCTCTCCAG 1589
Db 1739 AGGAGTATTTTCCCTCTTCTTACCTCTCTAAGGCTATGACCCCTCCACGCTCTCCAG 1798
QY 1590 CTGGGGAGTGGGGAGTGTATAGGAAAAGCCCCATCTCCCATCTGGGATAGGGAACCTTC 1649
Db 1799 CTGGGGAGTGGGGAGTGTATAGGAAAAGCCCCATCTCCCATCTGGGATAGGGAACCTTC 1858
QY 1650 CATCAGCTTTAACCTCTGGGAAATGCTGCTGCTCCCGCAGTACATCTTTGTTCTCTCCA 1709
Db 1859 CATCAGCTTTAACCTCTGGGAAATGCTGCTGCTCCCGCAGTACATCTTTGTTCTCTCCA 1918
QY 1710 CATCAGAGAGAGGAGTGGAGGAGGAGGAGTCTCAGTTAGCAGGAGTCCCGAGGACAA 1769
Db 1919 CATCAGAGAGAGGAGTGGAGGAGGAGGAGTCTCAGTTAGCAGGAGTCCCGAGGACAA 1978
QY 1770 GTACAGCT 1829
Db 1979 GTACAGCT 2038
QY 1830 ACCACTCTGGG-CCTTGGGGGGAGGAGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1888
Db 2039 ACCACTCTGGG-CCTTGGGGGGAGGAGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2098
QY 1889 GAACTCTATATCACAATTAAGAGAGAGTCTCAGACAGACNTCTGAGAGGCGGCTGCGAGAGT 1948
Db 2099 GAACTCTATATCACAATTAAGAGAGAGTCTCAGACAGACNTTAAGAGAGGCGGCTGCGAGT 2157
QY 1949 TCTCAACTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1983
Db 2158 TCTCAACTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2192

RESULT 4
LOCUS CO721116 2138 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 7050 from Patent WO02068579.
ACCESSION CO721116
VERSION CO721116.1 GI:4281973
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 7050 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1..2138
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 83.9%; Score 166.8; DB 6; Length 2138;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1813; Conservative 0; Mismatches 59; Indels 22; Gaps 9;

Db 381 CAACACATCAGAGTGCAGG--CTGAGAGCTGCGGCTGACGAGCACCGTCATGAACACC 438
QY 213 CACCTTAGACATACCAAGTCCAGAGCTGGCGAGAGAGTGAAGACCCGAGCGTGGTGA 272
Db 439 TTGGCT-----CAAGAGTGCAG-CTGCCGAGAGAGTGAAGCC--GACGTGGTGA 487
QY 273 CCTCGCTACTGTGCGGAGAGAGAGACAGCTCACTAGTCCGAGGACGCGTCAACCCGATGA 332
Db 488 CCGTGG---CTCTGGGAGAGAGCGGACGCTCACTATCCGGGACAGGGGTCAACCCGATGA 544
QY 333 CAGAACAGTCCAGAGAGCTTTCCGAGAGGAGATCAAAATTGAGTGGCTCTCCCTTCC 392
Db 545 CAGAACAGTCCAGAGAGCTTTCCGAGAGGAGATCAAAATTGAGTGGCTCTCCCTTCC 604
QY 393 TTTCCCTGCCCCCGCCCAAGCCTGACAGCAAAAGCTGCTAAACCCATTAGACAGC 452
Db 605 TTTCCCTGCCCCCGCCCAAGCCTGACAGCAAAAGCTGCTAAACCCATTAGACAGC 664
QY 453 TCCAGAGATCTCAGCCAGGTTCTTAGCCCTCAGAGCACCAGACCCAGTGAACATC 512
Db 665 TCCAGAGATCTCAGCCAGGTTCTTAGCCCTCAGAGCACCAGACCCAGTGAACATC 724
QY 513 CTCCCAACTAGAGGCGCTCCACTATCCAGAGGAGGAGGAGTCCCTGAGCTGACAC 572
Db 725 CTCCCAACTAGAGGCGCTCCACTATCCAGAGGAGGAGGAGTCCCTGAGCTGACAC 784
QY 573 ATGATGCCAGATTTCAATTTGGCTCCGTCCTTAATCCAGATACAGGGGCTGGAGT 632
Db 785 ATGATGCCAGATTTCAATTTGGCTCCGTCCTTAATCCAGATACAGGGGCTGGAGT 844
QY 633 CAGGAGAGAGATCTTAAGAGCCAGCTGGGTGAGGGAGTGGGACAGAGAGACATA 692
Db 845 CAGGAGAGAGATCTTAAGAGCCAGCTGGGTGAGGGAGTGGGAGAGAGAGACATA 904
QY 693 TGGGCAAGCTCTGACAGACAGACAGACAGACAAACCTCTGATCTATAGAGTCTTGACG 752
Db 905 TGGGCAAGCTCTGACAGACAGACAGACAGACAAACCTCTGATCTATAGAGTCTTGACG 964
QY 753 GGCAGAGGAGCAGAGGACCTGGAACCTCTTTGGCCAGAGGAGTGGAGAGACAGAGGA 812
Db 965 GGCAGAGGAGCAGAGGACCTGGAACCTCTTTGGCCAGAGGAGTGGAGAGACAGAGGA 1024
QY 813 AGGTCAAGGAGGAGGAGTCTATCTAAGTGAATTAATTTCCGAGGAGCTCAGAGAGGC 872
Db 1025 AGGTCAAGGAGGAGGAGTCTATCTAAGTGAATTAATTTCCGAGGAGGAGCTCAGAGAGGC 1084
QY 873 AAGAGAGAGAGGAGGAGTGAAGTAACTTCCCTCTTACAGAGCTCCAGGCCAGCCAGC 932
Db 1085 AAGAGAGAGAGGAGGAGTGAAGTAACTTCCCTCTTACAGAGCTCCAGGCCAGCCAGC 1144
QY 933 GAGCAGGAGTGGTCCGACCCGCTGCGCCAGCAGCTGGGCTGGTCCAGAGGAGAGCAT 992
Db 1145 GAGCAGGAGTGGTCCGACCCGCTGCGCCAGCAGCTGGGCTGGTCCAGAGGAGAGCAT 1204
QY 993 GGCACATCTGATATAGATGAGGATTTTCCATACAGCTGATTTGTAAGAACTGCATGA 1052
Db 1205 GGCACATCTGATATAGATGAGGATTTTCCATACAGCTGATTTGTAAGAACTGCATGA 1264
QY 1053 AACTCTGCGGCTGCTGCGCTCTGCGGCTCCAGAGGAGAGCATGAGGAGTGGGGGTG 1112
Db 1265 AACTCTGCGGCTGCTGCGGCTCTGCGGCTCCAGAGGAGAGCATGAGGAGTGGGGGTG 1324
QY 1325 GGGCTGTGCTTCTCCCTCCAGAGGCTGTGTTGTTGAGGAGTGGCTCCAGAGAGG 1384
Db 1173 ATACCTTAACAGATGAGAGCAGGAGCATGATGGGCTTTGGGTCTCGAGTTGAC 1232
QY 1385 ATACCTTAACAGAGAGTGAAGCAGGAGCATGATGGGCTTTGGGTCTCGAGTTGAC 1444
Db 1233 CCGAGCTTCTGACAGCTTCCCTCC-GGAGTCAAGCTCTCATCATCAGCTCTTTAA 1291
QY 1445 CCGAGCTTCTGACAGCTTCCCTCCGGGAGTCAAGCTCTCATCATCAGCTCTTTAA 1504

Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magatsuma, M., Shitatori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Suganara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Nishimura, K., Ishibashi, T., Yamashita, H., Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, Y., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, F., Hara, A., Sasaki, N., Aotaka, S., Yoshikawa, Y., Matsumura, H., Ichihara, T., Shiohama, N., Sano, S., Moriya, S., Nomiyama, H., Sato, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Oaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Maehuo, Y., Yamashita, R., Nakai, K., Tada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004)

14702039

2

Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hiroo, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Oaki, K., Sugiyama, T., Irie, K., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Murakami, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Maehuo, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project Unpublished 3 (bases 1 to 2142)

Isogai, T. and Yamamoto, J. Direct Submission Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3973, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1. 2142

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="cDNA"

/db_xref="taxon:9606"

/clone="HEART100185"

/cissue_type="heart"

/clone_lib="HEART1"

/note="Cloning vector: pME18SFL3"

ORIGIN

Query Match 83.6%; Score 1661; DB 9; Length 2142;

Best Local Similarity 95.6%; Pred. No. 0;

Matches 1809; Conservative 0; Mismatches 62; Indels 22; Gaps 9;

QY 33 GCCATCCCAACAGCCCGGTGGGCAAGCAATCAAGACCTAGAGAGCGCTATGA 92

Db 269 GGCCTCCCAAGCCCGCGGTGGGCAAGCAATCAAGACCTAGAGAGCGCTATGA 328

QY 93 GTTGCGGTGAGAGCTGAGAGACTTCACCTGAAAGCATATGTGCACCACTCCCAACA 152

Db 329 GTTGCGGTGAGAGCTGAGAGACTTCACCTGAAAGCATATGTGCACCACTCCA---A 385

QY 153 AACCAACATGAGGTGCGGGGCTGAGAAAGCTGCGGCTGAACGACACTGTACATGAAA 212

Db 386 CAACCAACATGAGGTGCGGG--CTGAGAAAGCTGCGGCTGAACGACACTGTACATGAAA 443

QY 213 CACCTTACGACTACAGAGGCGCAAGTGGCGGAGAGAGCGTGAACCCGAGCGTGGTGA 272

Db 444 TTGCGT-----CAAGAGCGAG--CTGCGGAGAGAGCGTGAACCC--GACGTGGTGA 492

QY 273 CCTCGGTACGTGTGAG 332

Db 493 CCTCGG---CTGCGGAG 549

QY 333 CAGAAACAGTTCAGCAACCTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 382

Db 550 CAGAAACAGTTCAGCAACCTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609

QY 393 TTTCCTGTCCTCCCG 452

Db 610 TTTCCTGTCCTCCCG 669

QY 453 TCAGAGACATCTAGCCCGAGCTTCTAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512

Db 670 TCAGAGACATCTAGCCCGAGCTTCTAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729

QY 513 CTCCCAACATGAG 572

Db 730 CTCCCAACATGAG 789

QY 573 ATGATGCCAGATTTTCAGATTTTGGCTCGCTCACTTATTCAGAGTACAGGGGCTGGGT 632

Db 790 ATGATGCCAGATTTTCAGATTTTGGCTCGCTCACTTATTCAGAGTACAGGGGCTGGGT 849

QY 633 CAGGAGAGAGATCTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 692

Db 850 CAGGAGAGAGATCTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 909

QY 693 TGGCAAGCTCTGACAG 752

Db 910 TGGCAAGCTCTGACAG 969

QY 753 GGCAGAGGAG 812

Db 970 GGCAGAGGAG 1029

QY 813 AGGTCAAG 872

Db 1030 AGGTCAAG 1089

QY 873 AAG 932

Db 1090 AAG 1149

QY 933 GACAG 992

Db 1150 GACAG 1209

QY 993 GCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGCTGTGTTGTGTAATAAATCTCATGA 1052

Db 1210 GCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGCTGTGTTGTGTAATAAATCTCATGA 1269

QY 1053 AACTCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 1112

Db 1270 AACTCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 1329

QY 1113 GGGCTGTCTTCTCCCTCCCAAGAGCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1172

Db 1330 GGGCTGTCTTCTCCCTCCCAAGAGCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1389

QY 1173 ATGACCTAAGAGAGATGGAAGCCAGGCGATGATGGGCGCTTGGTCCCTCGAGCTTGAC 1232
 DB 1390 ATGACCTAAGAGAGATGGAAGCCAGGCGATGATGGGCGCTTGGTCCCTCGAGCTTGAC 1449
 QY 1233 CCCAGCTTCTTGGCACCCTCCCTCC-CGGAGTCAGCTCTCAATCCATCCCTCTTTAA 1291
 DB 1450 CCCAGCTTCTTGGCACCCTCCCTCCGGGCGATGAGCTCTCAATCCATCCCTCTTTAA 1509
 QY 1292 TCTATGATCTATATGGCTCGGTGTGTATCA-ACAGACCCCTATGCTTGTCTTCAAT 1350
 DB 1510 TCTATGATCTATATGGCTCGGTGTGTATCAACACACACCCCTATGCTTGTCTTCAAT 1569
 QY 1351 ACTCAGCATTAACATGGTGTGAAGGCAATTCAGAGCTTCTCAATCAGATTACATC 1410
 DB 1570 ACTCAGCATTAACATGGTGTGAAGGCAATTCAGAGCTTCTCAATCAGATTACATC 1629
 QY 1411 TCCATTTTCATTAACGGGGAAACATCCCGAGCCACTGAGTGTGTGTCTTCACTGAA 1470
 DB 1630 TCCATTTTCATTAACGGGGAAACATCCCGAGCCACTGAGTGTGTGTCTTCACTGAA 1689
 QY 1471 GGTAGATCTGAACCCAGGGGTGTCACAGCTGTCTCACTCCCACTCTGTGGCACTGA 1530
 DB 1690 GGTAGATCTGAACCCAGGGGTGTCACAGCTGTCTCACTCCCACTCTGTGGCACTGA 1749
 QY 1531 GGAATATTTCCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCCACTTTCAG 1590
 DB 1750 GGAATATTTCCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCCACTTTCAG 1809
 QY 1591 TGGGGGATGGGGGAGTCATAGAGAAAGCCCCCATCTCCATCTGGGATAGGACCTTCC 1650
 DB 1810 TGGGGGATGGGGGAGTCATAGAGAAAGCCCCCATCTCCATCTGGGATAGGACCTTCC 1869
 QY 1651 ATCAGCTTAACTCCGGAATGCTGTGCTGCCCCAGTGACTCTGTTTCTCTCCAC 1710
 DB 1870 ATCAGCTTAACTCCGGAATGCTGTGCTGCCCCAGTGACTCTGTTTCTCTCCAC 1929
 QY 1711 ATACAGAGAGAGGATGAGAGGAGGATGGGCTCAGTTAGCAGGGGCTCCCAAGGCAAG 1770
 DB 1930 ATACAGAGAGAGGATGAGAGGAGGATGGGCTCAGTTAGCAGGGGCTCCCAAGGCAAG 1989
 QY 1771 TCAGCTCTCCCTCCATGCTCTGTGAGTGTGCTTAAAGGATGGGCTCTCACTCCCA 1830
 DB 1990 TCAGCTCTCCCTCCATGCTCTGTGAGTGTGCTTAAAGGATGGGCTCTCACTCCCA 2049
 QY 1831 CCACTCTGGG-CCTTGGGGAGAGCTGGGAGGGGCGTGGGAGAGCCCTTACGCTGG 1889
 DB 2050 CCACTCTGGGCGCTTGGGGAGAGAGCTGGGAGGGGCGTGGGAGAGCCCTTACGCTGG 2109
 QY 1890 AACCTGTATACATAAAGAGAGCTCTCAG 1922
 DB 2110 AACCTGTATACATAAAGAGAGCTCTCAG 2142

RESULT 7

AX747447 2410 bp mRNA linear PART 20-JUN-2003
 LOCUS AX747447
 DEFINITION Sequence 972 from Patent EP1308459.
 ACCESSION AX747447 GI:32131835
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens (human)

Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Taneichi, K. I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.
 Full-length cDNA sequences
 Patent: EP 1308459-A 972 07-MAY-2003;
 JOURNAL Helix Research Institute (JP) ; Research Association for Biotechnology (JP)

FEATURES
 source
 Location/Qualifiers
 1. 2410
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

Query Match 83.6%; Score 1660.6; DB 6; Length 2410;
 Best Local Similarity 95.6%; Pred. No. 0;
 Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

QY 33 GCCATCCCAACCAAGCCCGGTGGGAGGAGACATCAAGACCTTAGAGAGCGCTATGA 92
 DB 539 GGCCTTCCAGCCCGCCCGGTGGGAGGAGACATCAAGACCTTAGAGAGCGCTATGA 598
 QY 93 GTTGGGGTGAAGCTGAAGACTTCTCACTGAAAGCATATTGTCACTCACTCAACAA 152
 DB 599 GTTGGGGTGAAGCTGAAGACTTCTCACTGAAAGCATATTGTCACTCACTCAACAA 655
 QY 153 AACCAACATCGAGGTGGGAGGAGCTGAGAGGCTGAGAGGAGCACTGATCACTGAAA 212
 DB 656 CAACCAACATCGAGGTGGGAGGAGCTGAGAGGCTGAGAGGAGCACTGATCACTGAAA 713
 QY 213 CACCTTAAGACTAACAAGTGCAGACTGCCGAGAGAGCTGAGAGGAGCTGAGAGGAG 272
 DB 714 TTGGCT-----CAAAAGTGCAG-CTGGGAGAGAGGTGAGACCC--GACGTGGTGA 762
 QY 273 CCTGGGCTAAGTGGGAGAGAGGAGAGGAGCTGAGAGGAGCTGAGAGGAGCTGAGAGG 332
 DB 763 CTTGG--CTTGGGAGAGAGGAGAGGAGCTGAGAGGAGCTGAGAGGAGCTGAGAGG 819
 QY 333 CAGAAACATCGAGGAGAGCTTTCGAGAGGAGATCAAAATGAGTGGCTCTCCCTTCC 392
 DB 820 CAGAAACATCGAGGAGAGCTTTCGAGAGGAGATCAAAATGAGTGGCTCTCCCTTCC 879
 QY 393 TTTCCCTGTCCCCCGCCCGCCAGCTGAGAGGAGCTGAGAGGAGCTGAGAGGAGCT 452
 DB 880 TTTCCCTGTCCCCCGCCCGCCAGCTGAGAGGAGCTGAGAGGAGCTGAGAGGAGCT 939
 QY 453 TCCAGAGATCTCAAGCCAGGCTTTCGAGAGGAGCTGAGAGGAGCTGAGAGGAGCT 512
 DB 940 TCCAGAGATCTCAAGCCAGGCTTTCGAGAGGAGCTGAGAGGAGCTGAGAGGAGCT 999
 QY 513 CTCCTAACTAGGAGCTTCTCACTTATCAAGGAGGAGCTGAGAGGAGCTTCTGAGAG 572
 DB 1000 CTCCTAACTAGGAGCTTCTCACTTATCAAGGAGGAGCTGAGAGGAGCTTCTGAGAG 1059
 QY 573 ATGATGCCAGATTCAAGATTGGGCTCGTCACTTATCAAGAGTACAGGAGGCTGGAGT 632
 DB 1060 ATGATGCCAGATTCAAGATTGGGCTCGTCACTTATCAAGAGTACAGGAGGCTGGAGT 1119
 QY 633 CAGGAGAGAGATCTAAAGAACCCAGTGTGGGTCAAGGAGATGGGAGCTGAGAGGAG 692
 DB 1120 CAGGAGAGAGATCTAAAGAACCCAGTGTGGGTCAAGGAGATGGGAGCTGAGAGGAG 1179
 QY 693 TGGGAGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 752
 DB 1180 TGGGAGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
 QY 753 GGCAGAGGAG 812
 DB 1240 GGCAGAGGAG 1299
 QY 813 AGGTCAAG 872
 DB 1300 AGGTCAAG 1359
 QY 872 AAG 932
 DB 1360 AAG 1419
 QY 933 GAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 992

CDS

/clone_11b="NTONG2"
/note="cloning vector: pME18SFL3"
126..863
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC03846.1"
/db_xref="GI:21750828"
/translation="MGKQVNPFRPGNCGOLDREBGTQQRVTAKHTWSHCQVLMEDXDS
QWGHLDWVAHVAVSSQSGPQGGGIGATWQSGGLATWAGQWAPPL
RMASTLAQGDPEWEXKALSMEDPQSGFPMPSHSEPLAFAPRPGAGKIKLGGAYEFA
VDVPSPEPDIVTTSNNHTEVRAEKALAGCTWMTFAHKQLEPDVDPISVTSALRE
DGLITFARHPRHTEHVQTFRIEIKI"

ORIGIN

Query Match 83.6%; Score 1660.6; DB 9; Length 2410;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGGCAACATCAAGAACCCTAGAGAGCGCTATGA 92
Db 539 GGCCTTCCCAAGCCCGGTGGGAGGCAACATCAAGAACCCTAGAGAGCGCTATGA 598
QY 93 GTTGGCGTGAAGTGAAGACTCTGACCTGAAGACATCATTTGTACACCTCCAAACA 152
Db 599 GTTGGCGTGAAGTGAAGACTCTGACCTGAAGACATCATTTGTACACCTCCCA---A 655
QY 153 AACCAACATCGAGGTGGGGGCTGAGAAAGCTGGGCTGAAACGACACTGTACATGAAA 212
Db 656 CAACCAACATCGAGGTGGGGG--CTGAGAAAGCTGGGCTGACGGACCGTACATGAACACC 713
QY 213 CACCTTACGACTACACAAAGTCCAGACTGCGGAGAGAGCTGAGACCCCGAGCGTGGGTA 272
Db 714 TTGCGT-----CACAAAGTCCAG-CTGCGGAGAGAGTGGAGCC--GACGTGGTGA 762
QY 273 CCTGGCTACTGTGGGAGAGAGGACAGCTCACTAGTCCGGGACGCTACCCGATGA 332
Db 763 CCGTGG---CTGCGGAGAGAGAGGACGCTCACTAGTCCGGGACGCTACCCGATGA 819
QY 333 CAGAACGCTTCCAGCAACCTTCGGAGCGAGATCAAAATCTGAGTCCCTCTCCCTCC 392
Db 820 CAGAACGCTTCCAGCAACCTTCGGAGCGAGATCAAAATCTGAGTCCCTCTCCCTCC 879
QY 393 TTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAACCCTTACCAAGC 452
Db 880 TTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAACCCTTACCAAGC 919
QY 453 TCCAGACATCTCAAGCCAGGTTTACGCCCCCAGCGACCCCAAGCCAGGTGAGCATC 512
Db 940 TCCAGACATCTCAAGCCAGGTTTACGCCCCCAGCGACCCCAAGGTGAGCATC 999
QY 513 CTCCCAACTAGAGGCGCCCTCACTATCAAGGAGGAGCGAGGATCCCTGGGCTGAC 572
Db 1000 CTCCCAACTAGAGGCGCCCTCACTATCAAGGAGGAGCGAGGATCCCTGGGCTGAGC 1059
QY 573 ATGATGCCAATTTAGATTTGGCTTCGCTCACTTATCAAGTACAGGAGCTGGGGT 632
Db 1060 ATGATGCCAATTTAGATTTGGCTTCGCTCACTTATCAAGTACAGGAGCTGGGGT 1119
QY 633 CAGGGAAGGAAGATCTTAAAGAACCACTGAGGAGTCAAGGAGATGAGGACAGGACATA 692
Db 1120 CAGGGAAGGAAGATCTTAAAGAACCACTGAGGAGTCAAGGAGATGAGGACAGGACATA 1179
QY 693 TGAGCAAGCTCTGAGGAGACAGACAGACAAACCTCTGATCTATGAAGTCTGACAG 752
Db 1180 TGAGCAAGCTCTGAGGAGACAGACAGACAAACCTCTGATCTATGAAGTCTGACAG 1239
QY 753 GGCAGAGGAGACCAAGGAGCTTGAACCTCTTTGGCAGAGGAGTGGAGAGACAGAGGA 812
Db 1240 GGCAGAGGAGACCAAGGAGCTTGAACCTCTTTGGCAGAGGAGTGGAGAGACAGAGGA 1299
QY 813 AGGTCAAGGAGGAGGAGTGGCTATAGTGAATTAATTCGAGAGGCTCAGACAGGCC 872
Db 1300 AGGTCAAGGAGGAGGAGTGGCTATAGTGAATTAATTCGAGAGGCTCAGACAGGCC 1359

QY 873 AAGAGAGACAGCCCGTGAAGGTAACCTTCCCTTACAGAGCTCCAGGCCCAAGCCAGCAGC 932
Db 1360 AAGAGAGACAGCCCGTGAAGGTAACCTTCCCTTACAGAGCTCCAGGCCCAAGCCCAAGCCAGC 1419
QY 933 GAGCAGGCTGCTGCCACCCGTCGCCCGCCAGCCAGCTGAGTGGCCAGAGGACCAT 992
Db 1420 GAGCAGGCTGCTGCCACCCGTCGCCCGCCAGCCAGCTGAGTGGCCAGAGGACCAT 1479
QY 993 GCCAATCTGTATATAGATGGGGTTTTCGATATAGAGGAGTGGTGGTGAAGACATGACATGA 1052
Db 1480 GCCAATCTGTATATAGATGGGGTTTTCGATATAGAGGAGTGGTGGTGAAGACATGACATGA 1539
QY 1053 AACTCTGCGCTCTGCGCTGCTGAGGAGCTTCAGAGCAAGGCAAGTGGGGTGGGGTGG 1112
Db 1540 AACTCTGCGCTCTGCGCTGCTGAGGAGCTTCAGAGCAAGGCAAGTGGGGTGGGGTGG 1599
QY 1113 GGGCTGCTCTTCTCTCCCTCCCAAGGCTGTGTGTTGAGGGCTGCTCCCATGACAGAG 1172
Db 1600 GGGCTGCTCTTCTCTCCCTCCCAAGGCTGTGTGTTGAGGGCTGCTCCCATGACAGAG 1659
QY 1173 ATCACCATAACAGATGAGAGCCAGAGGATGATGAGGCTTTGGTCTCCAGGTTGAGC 1232
Db 1660 ATCACCATAACAGATGAGAGCCAGAGGATGATGAGGCTTTGGTCTCCAGGTTGAGC 1719
QY 1233 CCAGCTTCTTCCACCTTCCCTCC-GGAGTCAAGCTCTCCATCCATCCCTCTTTAA 1291
Db 1720 CCAGCTTCTTCCACCTTCCCTCC-GGAGTCAAGCTCTCCATCCATCCCTCTTTAA 1779
QY 1292 TGTATGAATCTATAGGCTGGGTGTGTATGA-AGACACCCCTATGCTTGTCTCAAT 1350
Db 1780 TGTATGAATCTATAGGCTGGGTGTGTATGA-AGACACCCCTATGCTTGTCTCAAT 1839
QY 1351 ACTCAGCATTAACATTTGTTAGGACCAATTCAGAGCTTTCTCAATCAGATTTACAAATC 1410
Db 1840 ACTCAGCATTAACATTTGTTAGGACCAATTCAGAGCTTTCTCAATCAGATTTACAAATC 1899
QY 1411 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCAGTGAAGTGTGCTTGTGCTGAA 1470
Db 1900 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCAGTGAAGTGTGCTTGTGCTGAA 1959
QY 1471 GGTATGATGTAACCCAGGAGTGTCAACAGCTGCTCACTCCCACTGGGACATGA 1530
Db 1960 GGTATGATGTAACCCAGGAGTGTCAACAGCTGCTCACTCCCACTGGGACATGA 2019
QY 1531 GGAATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACACCTCTCCACAGCTTTCAGC 1590
Db 2020 GGAATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACACCTCTCCACAGCTTTCAGC 2079
QY 1591 TGGGGGATGGGGGAGTCAATAGAAAGCCCATCTCCATCTGGAGATGAGGACCTTCC 1650
Db 2080 TGGGGGATGGGGGAGTCAATAGAAAGCCCATCTCCATCTGGAGATGAGGACCTTCC 2139
QY 1651 ATCAGCTTAACCTTGGAAATGCTGTGCTGCCAGTGAATCTTGTGTTGCTCCAC 1710
Db 2140 ATCAGCTTAACCTTGGAAATGCTGTGCTGCCAGTGAATCTTGTGTTGCTCCAC 2199
QY 1711 ATACAGAGCAGAGGTGAGAGGGGAAAGGTGGGTCTAGTTAGAGAGGGTCCCGAGGGCAAG 1770
Db 2200 ATACAGAGCAGAGGTGAGAGGGGAAAGGTGGGTCTAGTTAGAGAGGGTCCCGAGGGCAAG 2259
QY 1771 TCAGCTCTCCCTCATGCTCTCTGTGTCAGTGTGCTTAAAGGTGGCTCTCACTCCCA 1830
Db 2260 TCAGCTCTCCCTCATGCTCTCTGTGTCAGTGTGCTTAAAGGTGGCTCTCACTCCCA 2319
QY 1831 CCACTCTGGG--CCCTTGGGGAGAGCTGGGAGAGGGGGCGTGGAGACCCCTGAGCGCTGG 1889
Db 2320 CCACTCTGGG--CCCTTGGGGAGAGCTGGGAGAGGGGGCGTGGAGACCCCTGAGCGCTGG 2379
QY 1890 AACCTGTATACAAATTAAGGACAGTCTCAC 1920
Db 2380 AACCTGTATACAAATTAAGGACAGTCTCAC 2410

RESULT 9
 LOCUS CQ783843 2701 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 3983 from Patent EP1395643.
 ACCESSION CQ783843
 VERSION CQ783843.1 GI:45503747
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 REFERENCE
 AUTHORS Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Makamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 TITLE Primers for synthesizing full length cDNA clones and their use
 JOURNAL Patent: EP 1395643-A 3983 10-MAR-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source 1..2701
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 792..1151
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAF86888.1"
 /db_xref="GI:45503748"
 /translation="MOGLHSLITLAHPFAPRGAGNITKLGDVAVVDVDSPPED
 IIVTISNHI EVRAELADGTVMNTFAHKQLPEDVPTSVTSALRDSGLITLRAR
 HPHTHVQOTFRFEIKI"
 ORIGIN
 Query Match 83.5%; Score 1658.8; DB 6; Length 2701;
 Best Local Similarity 95.7%; Pred. No. 0;
 Matches 1806; Conservative 0; Mismatches 59; Indels 23; Gaps 9;
 QY CCCAACCAAGCCCCCGGTGGGCAAGCAACATCAAGACCTTAGAGAGCGCTATAGTTGC 98
 DB CACAGCCCGCCCGGTGGGCAAGCAACATCAAGACCTTAGAGAGCGCTATAGTTGC 892
 QY GGTGGACGTGAGAGATTTCTCACTGAAACATCATTTGACACCACTCCAAACCA 158
 DB GGTGGACGTGAGAGATTTCTCACTGAAACATCATTTGACACCACTCCAAACCA 949
 QY CATCGAGTGGGGGCTGAGAAAGCTGGCGCTGACGACACTATCATGAAACACTT 218
 DB CATCGAGTGGGGGCTGAGAAAGCTGGCGCTGACGACACTATCATGAAACACTT 1007
 QY 219 AGCACTACAAAGTCCCAACTGCGGAGAGGTGGACCCGCGGTGGGTGACTTGG 278
 DB 1008 -----CAAGATGTCAG-CTGCGGAGAGAGGTGGACCC--GAGTCTGGTGAACCTTGG 1056
 QY 279 CTACTGTGGGAG 338
 DB 1057 ----CTTGGGGAG 1113
 QY 339 AGCTTCAG 398
 DB 1114 AGCTTCAG 1173
 QY 399 TGTCCCCCGCCCGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
 DB 1174 TGTCCCCCGCCCGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1233
 QY 459 ACATTCAG 518
 DB 1234 ACATTCAG 1293
 QY 519 AACTAG 578
 DB 1294 AACTAG 1353

QY 579 CCNAGATTTCAGATTGGGCTCCGTCACTTAATCCAGAGTACAGGGGCTGGGGTCAAGGA 638
 DB 1354 CCNAGATTTCAGATTGGGCTCCGTCACTTAATCCAGAGTACAGGGGCTGGGGTCAAGGA 1413
 QY 639 AGAAGATCTTAAGAACCCACTGTGGGTGAGGGGAAATGGAGCAGAGACATATGGGC 698
 DB 1414 AGAAGATCTTAAGAACCCACTGTGGGTGAGGGGAAATGGAGCAGAGACATATGGGC 1473
 QY 699 AGCTTCAG 758
 DB 1474 AGCTTCAG 1533
 QY 759 GGGACCGAG 818
 DB 1534 GGGACCGAG 1593
 QY 819 CAGGCAAGGGTGCCTATCTAAGTGAATATTATGAGGGCTCAGCAAGGCCAAGAG 878
 DB 1594 CAGGCAAGGGTGCCTATCTAAGTGAATATTATGAGGGCTCAGCAAGGCCAAGAG 1653
 QY 879 AGACAGCGGTGAG 938
 DB 1654 AGACAGCGGTGAG 1713
 QY 939 GGTGCTGAG 998
 DB 1714 GGTGCTGAG 1773
 QY 999 TGTGTATATGATGGGTTTTCATATACAGCTGTTGTTGAAAACTGATGAATCC 1058
 DB 1774 TGTGTATATGATGGGTTTTCATATACAGCTGTTGTTGAAAACTGATGAATCC 1833
 QY 1059 TGGCGTCTGAG 1118
 DB 1834 TGGCGTCTGAG 1893
 QY 1119 GTCTTCTCCCTCCACAGGGCTGTGTCTGTTGAGGAGTGTCCATGACAGAGATCAC 1178
 DB 1894 GTCTTCTCCCTCCACAGGGCTGTGTCTGTTGAGGAGTGTCCATGACAGAGATCAC 1953
 QY 1179 TAAAGAGATGGAAGCAG 1238
 DB 1954 TAAAGAGATGGAAGCAG 2013
 QY 1239 TTCTTGCACCTTCCCTCC-GGCAGTCAAGCTCTCCATCCATCCCTCTTAATCTATG 1297
 DB 2014 TTCTTGCACCTTCCCTCCGGGAGAGTCAAGCTCTCCATCCATCCCTCTTAATCTATG 2073
 QY 1298 AATCTATAGGCTGGGTGTGTGA-ACACACCTTATGTTGTCTTCAATATCTAG 1356
 DB 2074 AATCTATAGGCTGGGTGTGTGAACACACACCTTATGTTGTCTTCAATATCTAG 2133
 QY 1357 CATTACATTTGTTGAGGAGCAATTCAGAGCTTTCAATCAGATTTCAGATTCGATTT 1416
 DB 2134 CATTACATTTGTTGAGGAGCAATTCAGAGCTTTCAATCAGATTTCAGATTCGATTT 2193
 QY 1417 TTCAATTAAGGGGAAATCCTCCGAGCAGCTGAGTGTGTTGTCACTGAAGTTAG 1476
 DB 2194 TTCAATTAAGGGGAAATCCTCCGAGCAGCTGAGTGTGTTGTCACTGAAGTTAG 2253
 QY 1477 ATCTGAACCGAGGAGTCAACGCTGCTCAACTCCCACTCTGGGAGAGAGAGAG 1536
 DB 2254 ATCTGAACCGAGGAGTCAACGCTGCTCAACTCCCACTCTGGGAGAGAGAGAG 2313
 QY 1537 TTTCCTCTCAATTTAATCTCTTAAGGCTATGACCCCTCCAGAGCTTTCCAGCTGGGG 1596
 DB 2314 TTTCCTCTCAATTTAATCTCTCTTAAGGCTATGACCCCTCCAGAGCTTTCCAGCTGGGG 2373
 QY 1597 ATGGGGAGAGTCAATAGAAAAGCCCAATCTCCATCTGGGATAGGAGACTTCCATCAG 1656
 DB 2374 ATGGGGAGAGTCAATAGAAAAGCCCAATCTCCATCTGGGATAGGAGACTTCCATCAG 2433

Accession	Sequence	Position
QY	CTTAAACCTGGGAAATGCTCTGCGCCCACTACTCTTGTTTCGTTCCCAATACAG	1716
Db	2434 CTTAAACCTGGGAAATGCTCTGCGCCCACTACTCTTGTTTCGTTCCCAATACAG	2493
QY	1717 AAGCAGGGTGGAGGGGAAAGGGTGGGTCTCAGTTAGCAGGGGTCCCAAGGCAATCAGCC	1776
Db	2494 AAGCAGGGTGGAGGGGAAAGGGTGGGTCTCAGTTAGCAGGGGTCCCAAGGCAATCAGCC	2553
QY	1777 TCTTCCCTCCATGAGCTCTCTGATCAGTATGAGCTTAGGGTGGGCTCTCACTCCCAACCACTC	1836
Db	2554 TCTTCCCTCCATGAGCTCTCTGATCAGTATGAGCTTAGGGTGGGCTCTCACTCCCAACCACTC	2613
QY	1837 TGGGCCC-TTGGGGGAGGAGCTGGGGAGGGGGCCGTTGGAGAGCCCTGACGCTTGAACCT	1894
Db	2614 TGGGCCCCCTTGGGGGAGGAGCTGGGGAGGGGGCCGTTGGAGAGCCCTGACGCTTGAACCT	2673
QY	1895 GTATACACAATAAAGAGACAGTCTCACAAG	1922
Db	2674 GTATACACAATAAAGAGACAGTCTCACAAG	2701

RESULT	10
BD127861	
LOCUS	BD127861
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.
ACCESSION	BD127861
VERSION	BD127861.1 GI:2322806
KEYWORDS	JP 2002017375-A;J292.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 2701)
AUTHORS
Oe, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
TITLE
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 3292 22-JAN-2002;

COMMENT

FN 06-20020173-5-4,3,292
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS (792)..(1148).

```

FEATURES
  source
    location/Qualifiers
      1..2701
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
```

Query Match	83.5%	Score 1658.8	DB 6	Length 2701
Best Local Similarity	95.7%	Pred. No. 0		
Matches 1806	Conservative	0	Mismatches 59	Indels 23
				Gaps 9

QY	39	CCCAACCAAGCCCCGGTGGGCAAGCAATCAAGCCCTACAGAGAGCGCTTGTAGTTGTC	98
Db	833	CACAGCCCCCGGTGGGGGCAAGCAATCAAGCCCTACAGAGAGCGCTTGTAGTTGTC	892
QY	99	GGTGGACGTGAGAGACTTCTCACTGGAAGACATCATTTGTACCACTTCAACCAACCAACCA	158
Db	893	GGTGGACGTGAGAGACTTCTCACTGGAAGACATCATTTGTACCACTTCAACCAACCAACCA	949

OY	159	TATGAGATGGGGGCTGAGAAACCTGAGGCTGAAAGGACACTGATCATGAACAACCTT	218
Db	950	CATGAGATGGGG--CTGAGAACTGGCGGCTGACGCGACCGTCATGAAACCTTGCT	1007
OY	219	ACGACTACACAGTCCAGACTGCGGAGACGTGACCCCGAGCGTGCGGTACCTGG	278
Db	1008	-----CACAAGTGCCAG--CTGCGGAGAGCGTGTGACCC--GACGTGCGTGCCTTGG	1056
OY	279	CTACTGTGGGAGGACGGAACGCTCTCACTAGTCCGGGACCGCTGACCCGATACAGAAC	338
Db	1057	---CTGTGCGGAGAGAGCGACGCTCTCACTATCCGGGACAGCGCTGACCCGATACAGAAC	1113
OY	339	ACGTCCAGCAGACCTTCCGACGGAGATCAAATCTGATGTCCTTCCCTTCC	398
Db	1114	ACGTCCAGCAGACCTTCCGACGGAGATCAAATCTGATGTCCTTCCCTTCC	1173
OY	399	TGTCCCCCGCCGACGCTGCGACGAAAGCTGCTAAACCCATTAACAAGCTCCAG	458
Db	1174	TGTCCCCCGCCGACGCTGCGACGAAAGCTTGTATCCCATTAACAAGCTCCAG	1233
OY	459	ACATCTAGCCCGAGTTCTAGCCCCCAACGCCAGACCCAGGTGGACCTCTCCA	518
Db	1234	ACATCTAGCCCGAGTTCTAGCCCCCAACGCCAGACCCAGGTGGACCTCTCCA	1293
OY	519	AACTAAGGACCTTCAACTCTATCCAGGGGACGGGACCTCCGTCCTGACACATGATG	578
Db	1294	AACTAAGGACCTTCAACTCTATCCAGGGGACGGGACCTCCGTCCTGACACATGATG	1355
OY	579	CCGAGATTTGAGATTTGGCCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGGTACAGGA	638
Db	1354	CCGAGATTTGAGATTTGGCCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGGTACAGGA	1413
OY	639	AGGAAGATCTTAAAGAACCACTGTGGGTCAAGGGAAATGGACACAGCAGACATATGGCA	698
Db	1414	AGGAAGATCTTAAAGAACCACTGTGGGTCAAGGGAAATGGACACAGCAGACATATGGCA	1477
OY	699	AGCTCTGAGGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTCTCAGGGCAAG	758
Db	1474	AGCTCTGAGGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTCTCAGGGCAAG	1533
OY	759	GGGACACAGGACCTTGGAACCTTCTTGGCCAAAGGGGATGGGAGACAGAGGAAAGTCA	818
Db	1534	GGGACACAGGACCTTGGAACCTTCTTGGCCAAAGGGGATGGGAGAGACAGAGGAAAGTCA	1593
OY	819	CAGACAAAGGGGCTACTATCTAAGTGAATTAATCCCGAGGGCTCAGCAAGGCCAAGAG	878
Db	1594	CAGGCAAGGGGCTACTATCTAAGTGAATTAATCCCGAGGGCTCAGCAAGGCCAAGAG	1653
OY	879	AGACACCGGTGACGGTAACTTCCCTCTACAGGCTTCAAGCCACGCCAGCGAGCAG	938
Db	1654	AGACACCGGTGACGGTAACTTCCCTCTACAGGCTTCAAGCCACGCCAGCGAGCAG	1713
OY	939	GCTGCTGCGCCACCCCGTGGCCCGACGAGCTGGGTGCGCAGGCGACAGCATCCACA	998
Db	1714	GCTGCTGCGCCACCCCGTGGCCCGACGAGCTGGGTGCGCAGGCGAGACCATCCACA	1777
OY	999	TCTGTATATAGATGGGTTTTTCCAAATACAGCTGGTGTGAAAACTGCATGAATCTC	1057
Db	1774	TCTGTATATAGATGGGTTTTTCCAAATACAGCTGGTGTGAAAACTGCATGAATCTC	1833
OY	1059	TGGCGTCTCGGCTGCTGCTGGGGGCTCCAGGACAAAGCCAAAGTGGGTTGGGGTGGGGCTG	1118
Db	1834	TGGCGTCTCGGCTGCTGCTGGGGGCTCCAGGACAAAGCCAAAGTGGGTTGGGGTGGGGCTG	1893
OY	1119	GTCCTTCTCCTCCCAAGGCTGTGTGTTGTTGCGGCTGTCTCCATGACAGACATCAC	1178
Db	1894	GTCCTTCTCCTCCCAAGGCTGTGTGTTGTTGCGGCTGTCTCCATGACAGACATCAC	1953
OY	1179	TAAACAGAGTGAAGCCAGGCAATGAGATGGGGCTTTGGGTCTCTGAGGTTTGAACCCACG	1238
Db	1954	TAAACAGAGTGAAGCCAGGCAATGAGATGGGGCTTTGGGTCTCTGAGGTTTGAACCCACG	2013

QY 1239 TTCTGGCACTTCCCTCC-GGCAGTCAGCTCTCATCATCCCTCTTTATCTATG 1297
 Db 2014 TTCTGGCACTTCCCTCCGGGAGTCAGCTCTCATCATCCCTCTTTATCTATG 2073
 QY 1298 ATCTATAGGCTGGTGTGTGTACA-ACAACCCCTATCTGTCTTCAATCTG 1356
 Db 2074 AATCTATAGGCTGGTGTGTGTACA-ACAACCCCTATCTGTCTTCAATCTG 2133
 QY 1357 CATTAACATTTGTTAGGCAATTCAGAGCTTCTCAATTCAGATTTCAATCTCCAT 1416
 Db 2134 CATTAACATTTGTTAGGCAATTCAGAGCTTCTCAATTCAGATTTCAATCTCCAT 2193
 QY 1417 TTCAATTAACGGGAAACATCCCGAGCACTGAGTGTGTCTTGTCTCACTGAAGTTAG 1476
 Db 2194 TTCAATTAACGGGAAACATCCCGAGCACTGAGTGTGTCTTGTCTCACTGAAGTTAG 2253
 QY 1477 ATCTGAACCCAGGGGTGTCAACNGCTGTCTCAACCTCCGACCTGTGGAGCAGAGTA 1536
 Db 2254 ATCTGAACCCAGGGGTGTCAACNGCTGTCTCAACCTCCGACCTGTGGAGCAGAGTA 2313
 QY 1537 TTTCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCGACGCTTTCAGCTGGAGG 1596
 Db 2314 TTTCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCGACGCTTTCAGCTGGAGG 2373
 QY 1597 ATGGGGGGAGTCTATAGGAAAGCCCATCTCCCATCTGGGATAGGACCTTCCATCAGC 1656
 Db 2374 ATGGGGGGAGTCTATAGGAAAGCCCATCTCCCATCTGGGATAGGACCTTCCATCAGC 2433
 QY 1657 CTTAACCTGTGGAAATAGCTGTGTGCTGCTCCGACGAGTCTTGTCTCCCACTAAGC 1716
 Db 2434 CTTAACCTGTGGAAATAGCTGTGTGCTGCTCCGACGAGTCTTGTCTCCCACTAAGC 2493
 QY 1717 AACGAGGCTGGAGGGGAGGAGGCTGTCAATTAGCAGGGGTCCCAAGGCAAGTCAAGC 1776
 Db 2494 AACGAGGCTGGAGGGGAGGAGGCTGTCAATTAGCAGGGGTCCCAAGGCAAGTCAAGC 2553
 QY 1777 TCCTCCCTCATCTCTCTGTGTCAGTGTGCTTAAAGGTGGCTCTCACTCCCACTCAGC 1836
 Db 2554 TCCTCCCTCATCTCTCTGTGTCAGTGTGCTTAAAGGTGGCTCTCACTCCCACTCAGC 2613
 QY 1837 TGGGCCCC-TTGGGGGAGAGCTGGGGAGGGGCCGTGGAGAGCCCTGAACCTGAACCT 1894
 Db 2614 TGGGCCCCCTTGGGGGAGAGAGCTGGGGAGGGGCCGTGGAGAGCCCTGAACCTGAACCT 2673
 QY 1895 GTATACACATAAAGAGAGTCTCAAG 1922
 Db 2674 GTATACACATAAAGAGAGTCTCAAG 2701
 RESULT 11
 AK075151 2701 bp mRNA linear PRI 03-SEP-2002
 LOCUS Homo sapiens CDNA FLJ90670 f.18, clone PLACB1005539, similar to
 DEFINITION ACTIN POLYMERIZATION INHIBITOR.
 ACCESSION AK075151
 VERSION AK075151.1 GI:22761054
 KEYWORDS oligo capping; f.18 (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1
 AUTHORS Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahari,K., Masuho,Y., Ono,T., Okano,K., Yoshikawa,Y.,
 Aotsubu,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Niimiya,K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2701)
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission

JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yam, Kisaazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO Human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing;
 Research Association for Biotechnology; cDNA library construction;
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
 sequencing and clone selection; Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 source
 1. .2701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACB1005539"
 /tissue="Placenta"
 /clone_id="PLACB1"
 /note="Cloning vector: pMB18FL3"
 ORIGIN
 Query Match 83.5%; Score 1658.8; DB 9; Length 2701;
 Best Local Similarity 95.7%; Pred. No. 0;
 Matches 1806; Conservative 0; Mismatches 59; Indels 23; Gaps 9;
 QY 39 CCACACCAAGCCCCGGTGGGAGGCAACATCAAGCCTTAGAGAGCCTTAGATTG 98
 Db 833 CACAGCCCGCCGGTGGGAGGCAACATCAAGCCTTAGAGAGCCTTAGATTG 892
 QY 99 GGTGAGCTGAGAGACTTCTCACTGAAGCATCATTTGTCACACCTCCAAACCA 158
 Db 893 GGTGAGCTGAGAGACTTCTCACTGAAGCATCATTTGTCACACCTCCAAACCA 949
 QY 159 CATGAGAGTGGGGGCTGAGAAAGCTGGGCTGAAAGGCACTGTATATGAACACTT 218
 Db 950 CATGAGAGTGGGG--CTGAGAAAGTGGGGCTGAGGAGCGGTATGAACACTTGGCT 1007
 QY 219 ACGACTACAAAGTGGCTGAGCTGCCGAGAGAGTGAACCCGAGCTGGGGTGA 278
 Db 1008 -----CACAAATGGCAG--CTGCCGAGAGAGTGAACCC--GACGTGGTGA 1056
 QY 279 CTACTGTGGGAGAGAGAGAGAGCTCACTAGTCCGGGCAAGGCTACCCGATACAGAA 338
 Db 1057 ---CTGTGGGGAGAGAGAGAGCTCACTATCCGGGCAAGGCTACCCGATACAGAA 1113
 QY 339 ACGTCCAGAGAGACTTCCGAGCGGAGATCAAAATCTGAGTGCCTTCCCTTTCC 398
 Db 1114 ACGTCCAGAGAGACTTCCGAGCGGAGATCAAAATCTGAGTGCCTTCCCTTTCC 1173
 QY 399 TGTCCCGCCGCCAGCGCTGCGAGCAAGAGCTGTAAACCCCATTAACAAGCTCCAG 458
 Db 1174 TGTCCCGCCGCCAGCGCTGCGAGCAAGAGCTGTAAACCCCATTAACAAGCTCCAG 1233
 QY 459 ACATCTCAGCGAGTTCTAGCCGCCAGAGCAGACCCAGCCAGGTGACATCTCCCA 518
 Db 1234 ACATCTCAGCGAGTTCTAGCCGCCAGAGCAGACCCAGCCAGGTGACATCTCCCA 1293
 QY 519 AACTAGGGCCCTCACTATATCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 578
 Db 1294 AACTAGGGCCCTCACTATATCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1353
 QY 579 CCCAGATTTCAGATTGGCTCGTCACTTAATTCAGAGTACAGGGGCTGGGCTAGGGA 638
 Db 1354 CCCAGATTTCAGATTGGCTCGTCACTTAATTCAGAGTACAGGGGCTGGGCTAGGGA 1413
 QY 639 AGGAAGATTTAAGAACCCAGCTGTGGGTAGGGGAATGGGACAGAGACATATGGGCA 698
 Db 1414 AGGAAGATTTAAGAACCCAGCTGTGGGTAGGGGAATGGGACAGAGACATATGGGCA 1473
 QY 699 AGCTTCAGAGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTTCAGAGGCAAG 758
 Db 1474 AGCTTCAGAGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTTCAGAGGCAAG 1533

QY 759 GGGACCAAGGAGCTGGAAACCTCTTGGCCAAAGGAGATGGGAGAGACAGAGGAGAAAGTCA 818
 Db 1534 GGGACCAAGGAGCTGGAAACCTCTTGGCCAAAGGAGATGGGAGAGACAGAGGAGAAAGTCA 1593
 QY 819 CAGGCAAGGAGCTGATCTAAGTGAACCTAATGGCCGAGGAGCTCAGCAAGGCGCAAGAG 878
 Db 1594 CAGGCAAGGAGCTGATCTAAGTGAACCTAATGGCCGAGGAGCTCAGCAAGGCGCAAGAG 1653
 QY 879 AGACAGCTGACGATTAACCTTCCCTCTAAGCCTCCAGCCCAAGCCCAAGCAGAGCAG 938
 Db 1654 AGACAGCTGACGATTAACCTTCCCTCTAAGCCTCCAGCCCAAGCCCAAGCAGAGCAG 1713
 QY 939 GCTGCTGCCCCACCCCGTGGCCCAAGCAGCTGCTGTCAGAGGAGAGCCATGCGACA 998
 Db 1714 GCTGCTGCCCCACCCCGTGGCCCAAGCAGCTGCTGTCAGAGGAGAGCCATGCGACA 1773
 QY 999 TCTGTATATAGATGGGGTTTTTCCAAATACAGCTGTTCTGTGAAAAAATGCAATGAACTCC 1058
 Db 1774 TCTGTATATAGATGGGGTTTTTCCAAATACAGCTGTTCTGTGAAAAAATGCAATGAACTCC 1833
 QY 1059 TGGCGTCTGCGCTGCTGTCGAGGCTCCAGGCAAGGCGCAAGTGGGGGTTGGGGGCTG 1118
 Db 1834 TGGCGTCTGCGCTGCTGTCGAGGCTCCAGGCAAGGCGCAAGTGGGGGTTGGGGGCTG 1893
 QY 1119 GTCCCTTCTCCCTCCCAAGGCTGTGTTCTGAGGAGCTGCTCCATGCAAGCAGATCACC 1178
 Db 1894 GTCCCTTCTCCCTCCCAAGGCTGTGTTCTGAGGAGCTGCTCCATGCAAGCAGATCACC 1953
 QY 1179 TAAACAGATGGAAGCCAGGCAATGATGAGGAGCTTTGGGCTCTGAGGTTGAGCCCAAG 1238
 Db 1954 TAAACAGATGGAAGCCAGGCAATGATGAGGAGCTTTGGGCTCTGAGGTTGAGCCCAAG 2013
 QY 1239 TTTCTGACACCTTCCCTCC-AGCAGTCACTGCAATGCAATGCAATGCAATGCAATGCAATG 1297
 Db 2014 TTTCTGACACCTTCCCTCC-AGCAGTCACTGCAATGCAATGCAATGCAATGCAATGCAATG 2073
 QY 1298 AATCTATAGCTGGTGTGTGTGAACA-ACAACCCCTATGCTGTCTTCAATGACTCAG 1356
 Db 2074 AATCTATAGCTGGTGTGTGTGAACAACAACCCCTATGCTGTCTTCAATGACTCAG 2133
 QY 1357 CATTAACATGTTGTTGAGAGGCAAAATTCAGACCTTCTCAATTCAGATTAAATCTCAAT 1416
 Db 2134 CATTAACATGTTGTTGAGAGGCAAAATTCAGACCTTCTCAATTCAGATTAAATCTCAAT 2193
 QY 1417 TTTCTAATACGGGAGAAATCCCGAGGCACTGAGTGTGCTGCTTGTCACTGAGAGTTAG 1476
 Db 2194 TTTCTAATACGGGAGAAATCCCGAGGCACTGAGTGTGCTGCTTGTCACTGAGAGTTAG 2253
 QY 1477 ATCTGAACCAAGGAGTGTCAACNGTCTCTCAACTCCCACTCTGTGGGCACTGAGAGTGA 1536
 Db 2254 ATCTGAACCAAGGAGTGTCAACNGTCTCTCAACTCCCACTCTGTGGGCACTGAGAGTGA 2213
 QY 1537 TTTCCCTCATTTCACTCTCTAAGGATGAGCAACCTCCCGAGTCTTCAGTGGGG 1596
 Db 2314 TTTCCCTCATTTCACTCTCTAAGGATGAGCAACCTCCCGAGTCTTCAGTGGGG 2373
 QY 1597 ATGGGGGAGTCAATGAGAAAAAGCCCGCATCTCCCATGAGATGAGGAGCTTCATCAGC 1656
 Db 2374 ATGGGGGAGTCAATGAGAAAAAGCCCGCATCTCCCATGAGATGAGGAGCTTCATCAGC 2433
 QY 1657 CTTAAACCTGGGAAATGCTGCTGAGCCCGAGTCACTTGGTTTCTGCTCCCAATACAG 1716
 Db 2434 CTTAAACCTGGGAAATGCTGCTGAGCCCGAGTCACTTGGTTTCTGCTCCCAATACAG 2493
 QY 1717 AAGCAGGATGAGGAGGAGGATGGGTCTCAATTAGCAGGAGTCCCAAGGCAAGTCAAGC 1776
 Db 2494 AAGCAGGATGAGGAGGAGGATGGGTCTCAATTAGCAGGAGTCCCAAGGCAAGTCAAGC 2553
 QY 1777 TCTTCCCTCCATGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1836
 Db 2554 TCTTCCCTCCATGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2613
 QY 1837 TGGGCCC--TTGGGGGAGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1894

Db 2614 TGGGCCCCCTGGGGGAGAGACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2673
 QY 1895 GTATACACATATAAGACAGTCTCAG 1922
 Db 2674 GTATACACATATAAGACAGTCTCAG 2701
 RESULT 12
 HSM803488 2747 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens mRNA; CDNA DKFZp686p1416 (from clone DKFZp686p1416).
 DEFINITION
 ACCESSION AL832181
 VERSION AL832181.1 GI:21732726
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 2747)
 AUTHORS Manbut, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Oeanger, A.,
 Pobo, G., Han, M. and Wiemann, S.
 DIRECT SUBMISSION
 TITLE Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
 JOURNAL
 COMMENT
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGORA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 This clone (DKFZp686p1416) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://mips.gsf.de/proj/cdna/>.
 FEATURES
 source
 1..2747
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686p1416"
 /tissue_type="human cervix"
 /clone_id="686 (synonym: hicc). Vector pSport1_Sfi; host
 DH10B; sites SfiIa + SfiIb"
 /dev_stage="adult"
 2626..2631
 2641
 polyA_signal
 polyA_site
 ORIGIN
 Query Match 83.3%; Score 1656; DB 9; Length 2747;
 Best Local Similarity 95.6%; Pred. No. 0;
 Matches 1804; Conservative 0; Mismatches 62; Indels 22; Gaps 9;
 QY 39 CCCAACCAAGCCCGGATGGGAGAGCAATCAAGACCTAGAGAGCGCTAATGATTGGC 98
 Db 778 CACAGCCCGCCCGGATGGGAGAGCAATCAAGACCTAGAGAGCGCTAATGATTGGC 837
 QY 99 GGTGACGTGAGAGACTTCTCACTGAAAGCATCTTGTCACTCACTTCCAAACCA 158
 Db 838 GGTGACGTGAGAGACTTCTCACTGAAAGCATCTTGTCACTCACTTCCAAACCA 894
 QY 159 CATCGAGGAGGAGGAGGAGTGAAGAGCTGGGAGTGAAGAGCACTGATCAATGAAACACTT 218
 Db 895 CATCGAGGAGGAGGAGGAGTGAAGAGCTGGGAGTGAAGAGCACTGATCAATGAAACACTT 952
 QY 219 AGCACTACCAAGTSCCAAGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 278
 Db 953 -----CACAGTCCAG-CTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1001
 QY 279 CTACTGTGGGAG 338
 Db 1002 ---CTGTGGGAG 1058
 QY 339 AGTCACAGAGACCTTCGAG 398

Db	1059	ACGTCCAGCAGACTTTCGGGACGGAGATCGAATCTGATGCTCTCCCTTCCCTTTTCC	1118
OY	399	TGTCCCCCGCCCAACGCTTCGACAAAGCCTGCTAACCCATTACAGAGCTCCAGG	458
Db	1119	TGTCCCCCGCCCAACGCTTCGACAAAGCCTGCTAACCCATTACAGAGCTCCAGG	1178
OY	459	ACATCTCAGCCAGGTTGTAGCCCCCAGCAGCACCAGACGCCAGGTGAGCCATCCTCCCA	516
Db	1179	ACATCTCAGCCAGGTTGTAGCCCCCAGCAGCACCAGACGCCAGGTGAGCCATCCTCCCA	1238
OY	519	AACTAGGGCCCTCCACTCTATCCAGGGCAGGGCCTCCCTGCTGACATGATG	578
Db	1239	AACTAGGGCCCTCCACTCTATCCAGGGCAGGGCCTCCCTGCTGACATGATG	1298
OY	579	CCCAATTTCCAGATTTCGCTTCGCACTTAATCCAGATCAGGGGCTGGGGTCAAGGA	638
Db	1299	CCCAATTTCCAGATTTCGCTTCGCACTTAATCCAGATCAGGGGCTGGGGTCAAGGA	1358
OY	639	AGGAGATCTTAAAGAACCCACTGTGGTTCAGGGGAAATGGACACAGCATATGGGCA	698
Db	1359	AGGAGATCTTAAAGAACCCACTGTGTGTTCAGGGGATGGACACAGCATATGGGCA	1418
OY	699	AGCTTCGACGACAGACAGACAAACCTTCGATCTATGATGCTTCGACAGGGCAAG	758
Db	1419	AGCTTCGACGACAGACAGACAAACCTTCGATCTATGATGCTTCGACAGGGCAAG	1478
OY	759	GGGACCCAGGAGCCTTGAACCTCTTGGCCAGGGGAGTGGAGAGACAGAGGAAAGTCA	818
Db	1479	GGGACCCAGGAGCCTTGAACCTCTTGGCCAGGGGAGTGGAGAGAGCAAGGAAAGTCA	1538
OY	819	CAGGCAAGGATGCTATCTAATGAGAACTTAATTGCCAGAGGCTCAGCAAGGCCAAGAG	878
Db	1539	CAGGCAAGGATGCTATCTAATGAGAACTTAATTGCCAGAGGCTCAGCAAGGCCAAGAG	1598
OY	879	AGACAGCCGTGACGGTAACTTCCCTTTACCAAGCTTCAAGCCCCACGACGAGAGAG	938
Db	1599	AGACAGCCGTGACGGTAACTTCCCTTTACCAAGCTTCAAGCCCCACGACGAGAGAG	1658
OY	939	GCTGCTCCACACCCCTGACCCCAAGCAGCTGTGTGACCAAGGACAGAGCATGACACA	998
Db	1659	GCTGCTCCACACCCCTGACCCCAAGCAGCTGTGTGACCAAGGACAGAGCATGACACA	1718
OY	999	TCTGTATATAGATGGGGTTTTTCCAATACAGCTGGTTGTGAAAACCTGCATGAATCC	1058
Db	1719	TCTGTATATAGATGGGGTTTTTCCAATACAGCTGGTTGTGAAAACCTGCATGAATCC	1778
OY	1059	TGCGGTCTGGACCTGCTGGGGCCCTCCAGGCAAGGCAAGGGGCTTGGGGTGGGGCTG	1118
Db	1779	TGCGGTCTGGACCTGCTGGGGCCCTCCAGGCAAGGCAAGGGGCTTGGGGTGGGGCTG	1838
OY	1119	GTCCTTCCCTCCCAAGCCTGTGTCTGTTGAGGGAGCTGCCCATGACAGAGATACC	1178
Db	1839	GTCCTTCCCTCCCAAGCCTGTGTGTCTTGGGGGTGTGCTCCANTGACAGAGATACC	1898
OY	1179	TAAACAGATGAAGCCAGGGCATAGATGGGGCTTTGGGTCTCGAGGTTGACCCAGC	1238
Db	1899	TAAACAGATGAAGCCAGGGCATAGATGGGGCTTTGGGTCTCGAGGTTGACCCAGC	1958
OY	1239	TTCTTGCCACCTTCCCTCC-CGACAGTACGTCCACATCCCTCCTTATATCTATG	1297
Db	1359	TTCTTGCCACCTTCCCTCCCGGAGTCAAGTCTCCATCATCCCTCCTTATATCTATG	2018
OY	1298	AATCTATAGCTGAGTGTGTATACA-AACACCCCTATGTGTCTTCAATATCTCAG	1356
Db	2019	AATCTATAGCTGAGTGTGTGTATACAACACCCCTATGTGTCTTCAATATCTCAG	2078
OY	1357	CATATACATTTGGTGAAGCCAAATTCAGAGCTTCTCAAAATCAATTTACATTCATAT	1416
Db	2079	CATATACATTTGGTGAAGCCAAATTCAGAGCTTCTCAAAATCAATTTACATTCATAT	2138
OY	1417	TTCAATTACGGGGAAACATCCCGAGCCATGAGTGTGTGTGCTTGTCACTGAAGGTAG	1476

Db	2139	TTCAATTAAAGGGGAAAAACATCCCCGAGACACTGAGTGTCTGTCTTGTCACTGAAAGGTTGG	21398
OY	1477	ATCTGAACCCAGGGGTGTCAACNCGTGTCTCTCACTCCCACTCTTGGGCACTGAGAGTA	15366
Db	2199	ATCTGAACCCAGGGGTGTCAACAGGTGTCTCTCACTCCCACTCTTGGGCACTGAGAGTA	22588
OY	1537	TTTCCCCCTCATCTACCGCTCTAAGGGCATGACCCCTCCACAGCTCTCCAGGTGGGGG	15366
Db	2259	TTTCCCCCTCATCTTCACTCTCTTAAGGCTATGACCCCTCCCAAGTCTTCCAGGTGGGGG	23188
OY	1597	ATGCGGGGAGTCATAGGAAAAGCCCCCATCTCCCATCTGGGATAGGAGCCTTCATCAGC	16556
Db	2319	ATGCGGGGAGTCATAGGAAAAGCCCCCATCTCCCATCTGGGATAGGAGACCTTCATCAGC	23788
OY	1657	CTTAACCTCGGGAATGCGTGCCTGCCCCAGTAGACTCTGGTTTCGTCTCCCAATACAG	17176
Db	2379	CTTAACCTCGGGAATGCGTGCCTGCCCCAGTAGACTCTGGTTTCGTCTCCCAATACAG	24388
OY	1717	AAGCAGGAGTGAGAGGGGAAGGGGTGGGTCTCAATTGACAGGGGTCCCAAGGCAATCAGCC	17766
Db	2439	AAGCAGGAGTGAGAGGGGAAGGGGTGGGTCTCAATTGACAGGGGTCCCAAGGCAATCAGCC	24988
OY	1777	TTCCTCCCTCCATGCTCTCTTGTGTAGTGTGCTTAAGGTGGCTCTCACTCCCACTC	18366
Db	2499	TTCCTCCCTCCATGCTCTCTTGTGTAGTGTGCTTAAGGTGGCTCTCACTCCCACTC	25588
OY	1837	TGGG-CCCTTGGGGAGAGACTGGGGAGAGGGGCCGTGGAGAGCCCTGACGCTGAAACCTG	18956
Db	2559	TGGGCCCCCTTGGGGAGAGACTGGGGAGAGGGGCCGTGGAGAGCCCTGACGCTGAAACCTG	26188
OY	1896	TATACACAAATTAAGGACAGTCTCAAGA	1923
Db	2619	TATACACAAATTAAGGACAGTCTCAAAA	2646

RESULT	13
HSM806124	
LOCUS	
DEFINITION	HSM806124 Homo sapiens mRNA; CDNA DKFZ686A14192 (from clone DKFZ686A14192);
ACCESSION	BX537890 complete cds.
VERSION	BX537890.1 GI:31874066
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 2163)
Poustka, A., Albert, R., Moosmayer, P., Schnupp, I., Wellenreuther, R.,
Mewes, H. W., Weill, B., Amlid, C., Osanger, A., Fodor, G., Han, M. and
Wiemann, S.

TITLE	Direct Submission
JOURNAL	Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT	Clone from S. Wienann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the CDNA sequencing consortium of the German Genome Project. This clone (DKFZ866A14192) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonezrzd@rzpd.de. Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cdna/>.

```

FEATURES
    source
        location/Qualifiers
            1..2163
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /map="1p36.13"
                /clone="DKFZp686A14192"
                /tissue_type="human esophagus tumor"
                /clone_id="686 (SYNORMY: h1ec3). Vector dsport1 Sfi: host

```

gene DH10B, sites SfilA + SfilB
/dev stage="adult"
1. 2163
/gene="DKFZp686A14192"
<1. 577
/note="DKFZp686A14192"
/note="Heat-shock protein, beta-7, differentially spliced"
/codon start=2
/product="hypothetical protein"
/protein_id="CAD97949.1"
/db_xref="GI:31874067"
/translation="ARPEACATLSCPAFARLGRARRRRPMSHRTSSTRARRSRHS
SSSSSSSTSSSARLADPPEKALSWFSDFGFMSHPSEPLAFPRGAGNLIK
TLGDAYEFAVDYDFSPEDIVTTSNNHTELADGTVMNTFAHKCOLPEDVPTSVTS
ALREDSLTTRRRHPTHEVQOTFRTEIKI"
2107. 2112
/gene="DKFZp686A14192"
2122
/gene="DKFZp686A14192"

polYA_signal
polYA_site

ORIGIN

Query Match 82.9%; Score 1646.8; DB 9; Length 2163;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 59; Indels 37; Gaps 8;

33 GCCCATCCCAACCAAGCCCGGTGGGCGAGCAATCATAGACCCCTTGGAGAGCCCTATGA 92
268 GGCCTTCCAGCCCGCCCGGTGGGCGAGCAATCATAGACCCCTTGGAGAGCCCTATGA 327
93 GTTGGGCTGGAGGTGAGACCTTCTACCTGAAGCATGATGTGACCAACCTCCAAACA 152
328 GTTGGGCTGGAGGTGAGACCTTCTACCTGAAGCATGATGTGACCAACCTCCAAACA 387
153 AACCAACATGAGGTGGGCGGTGAGAAAGCTGGCGCTGAACGACATCTAATGAA 212
388 CCACATC-----GAGCTGGCGGTGAGCGGACCGTCATGAACAC 427
213 CACCTTACGACTACACAAAGTCCAGAGTGGCGAGAGAGTGGAGCCCGAGCGGTGA 272
428 TTGGCT-----CACAAATGCCAG-CTGCCGAGAGCGTGAACC--GACGTGAGTA 476
273 CCTGGCTACTGTGGGAGAGCGGACGCTCTAAGTCCGGGACGCGTCAACCCGATA 332
477 CCTGG---CTGGCGGAGAGAGCGGACGCTCTAAGTCCGGGACGCGTCAACCCGATA 533
333 CAGAACGCTTCAGACAGACCTTCCGAGCGAGATCAAAATCTGAGTGCCTCTCCCTCC 392
534 CAGAACGCTTCAGACAGACCTTCCGAGCGAGATCAAAATCTGAGTGCCTCTCCCTCC 593
393 TTTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAAGCCCTTCAACAGAG 452
594 TTTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAAGCCCTTCAACAGAG 653
453 TCCAGACATCTCAGCCAGGTTCTAGCCGCCAGAGACCCAGACCCCAAGTGAACATC 512
654 TCCAGACATCTCAGCCAGGTTCTAGCCGCCAGAGACCCCAAGTGAACATC 713
513 CTCCCAACTAGGCGCTTCACTCTATCCAGGGAGGCGAGGACTCCCTGGGCTGAC 572
714 CTCCCAACTAGGCGCTTCACTCTATCCAGGGAGGCGAGGACTCCCTGGGCTGAC 773
573 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGTGGGGT 833
774 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGTGGGGT 893
833 CAGGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGAGATGGGACGAGGAGATA 893
834 CAGGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGAGATGGGACGAGGAGATA 893
693 TGGCAAGCTCTGAGGAGACAGACAGACAAACCTCTGATTAAGTCTTGAG 752
894 TGGCAAGCTCTGAGGAGACAGACAGACAAACCTCTGATTAAGTCTTGAG 953

753 GGCAGAGGAGACCAAGGACCTTGAACCTCTTGGCCAAAGGAGTGGAGAGACAGAGGA 812
954 GGCAGAGGAGACCAAGGACCTTGAACCTCTTGGCCAAAGGAGTGGAGAGACAGAGGA 1013
813 AGGTACAGGAGCAAGGAGTCCCTATCTAAGTGAATTAATGGCCGAGGAGTCAAGAGGC 872
1014 AGGTACAGGAGCAAGGAGTCCCTATCTAAGTGAATTAATGGCCGAGGAGTCAAGAGGC 1073
873 AAGAGAGACACCGGTGACGCTTAACTTCCCTTACAGGCTCCCAAGCCCAAGCAGC 932
1074 AAGAGAGACACCGGTGACGCTTAACTTCCCTTACAGGCTCCCAAGCCCAAGCAGC 1133
933 GAGCAGGCTGCTGCCACCCCGTGGCCCGCCAGCCAGCAGTGTGTGCCAGGCAAGCAT 992
1134 GAGCAGGCTGCTGCCACCCCGTGGCCCGCCAGCCAGCAGTGTGTGCCAGGCAAGCAT 1193
993 GCCACATCTGATATAGATGGGGTTTTTCCATACAGCTGGTGTGTAATACTGATGA 1052
1194 GCCACATCTGATATAGATGGGGTTTTTCCATACAGCTGGTGTGATTAACCTGATGA 1253
1053 AACTCTGCGCTGCTGCGCTGCTGGGCGCTTCCAGGACCAAGGCAAGTGGGTTGGGCTG 1112
1254 AACTCTGCGCTGCTGCGCTGCTGGGCGCTTCCAGGACCAAGGCAAGTGGGTTGGGCTG 1313
1113 GGGCTGTGCTTCTTCCCTCCCAAGGCGCTGTGTTCTTGGGCTGTGCCATGCAAGACAG 1172
1314 GGGCTGTGCTTCTTCCCTCCCAAGGCGCTGTGTTCTTGGGCTGTGCCATGCAAGACAG 1373
1173 ATCACCTAACAGAGATGGAAGCCAGGAGCATGATGGGCTTGGGCTCTGAGGTTGAC 1232
1374 ATCACCTAACAGAGATGGAAGCCAGGAGCATGATGGGCTTGGGCTCTGAGGTTGAC 1433
1233 CCGAGCTTCTTCCACCTTCCCTTCC--GGAGTACGCTCTGATTCATACCTCCCTTTAA 1291
1434 CCGAGCTTCTTCCACCTTCCCTTCCCGGAGTACGCTCTGATTCATACCTCCCTTTAA 1493
1292 TCTATGAATCTAAGGCTCGGTTGTGTAAAC-ACACACCCCTATCGTTGCTTCAAT 1350
1494 TCTATGAATCTAAGGCTCGGTTGTGTAAACACACACCCCTATCGTTGCTTCAAT 1553
1351 ACTCAGCATTAACATGTTGTTGAGGCGCAATTCAGAGCTTCTCAATCAGATTTCATC 1410
1554 ACTCAGCATTAACATGTTGTTGAGGCGCAATTCAGAGCTTCTCAATCAGATTTCATC 1613
1411 TCCATTTTCATTAAGGAGAAACATCCCGAGCCCTGAGTGTGCTTGTGCTGAA 1470
1614 TCCATTTTCATTAAGGAGAAACATCCCGAGCCCTGAGTGTGCTTGTGCTGAA 1673
1471 GGTAGATCTGAACCCAGGAGTGTCAACNGCTGCTCTCAACTCCCACTTGGGACATGA 1530
1674 GGTAGATCTGAACCCAGGAGTGTCAACNGCTGCTCTCAACTCCCACTTGGGACATGA 1733
1531 GAGATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 1590
1734 GAGATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 1793
1591 TGGGGGATGGGGGAGTGAAGAGAAAGCCCATCTCCCATCTGGGATAGGAGACTTCC 1650
1794 TGGGGGATGGGGGAGTGAAGAGAAAGCCCATCTCCCATCTGGGATAGGAGACTTCC 1853
1651 ATCAGCTTAAACCTCTGGGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1710
1854 ATCAGCTTAAACCTCTGGGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913
1711 ATCAGAGACAGAGTGAAGGAGGAGGAGTGAAGGAGTGAAGGAGGAGTGAAGGAGGAG 1770
1914 ATCAGAGACAGAGTGAAGGAGGAGGAGTGAAGGAGTGAAGGAGGAGTGAAGGAGGAG 1973
1771 TGAAGCTCTCTCTCAATGCTCTTGTGCTGAGTGTGCTTGAAGTGTGCTCTCACTCCA 1830
1974 TGAAGCTCTCTCTCAATGCTCTTGTGCTGAGTGTGCTTGAAGTGTGCTCTCACTCCA 2033
1831 CCACTCTGGG-CCCTTGGGAGAGACTGGGAGAGGAGGCGGTGGAGAGAGCTGACGTGG 1889

Db 2034 CCACTCTGGGCCCCCTTGGGGAGAGACTGGGAGGGGCGCTGGAGAGAGCCCTGACGCTGG 2093

Qy 1890 AACCTGTATACACATTAAGAGACAGTCTCAGAGA 1923

Db 2094 AACCTGTATACACATTAAGAGACAGTCTCAGAGA 2127

RESULT 14

AF155908 2152 bp mRNA linear PRI 27-DEC-1999

LOCUS AF155908

DEFINITION Homo sapiens cardiovascular heat shock protein mRNA, complete cds.

ACCESSION AF155908

VERSION AF155908.1 GI:6635998

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 2152)

REFERENCE Kriebel, S., Faivre, J.F., Robert, P., Le Douarin, B., Brunet-Latignon, N., Leftere, I., Bouzyk, M.M., Anderson, K.M., Grottel, L.D., Tobin, F.L., Souchet, M. and Brill, A. Identification and characterization of cHSP. A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues

JOURNAL J. Biol. Chem. 274 (51), 36592-36600 (1999)

MEDLINE 2062883

PUBMED 10593960

REFERENCE 2 (bases 1 to 2152)

AUTHORS Kriebel, S., Faivre, J.F., Robert, P., Le Douarin, B., Brunet-Latignon, N., Leftere, I., Bouzyk, M.M., Anderson, K.A., Grottel, L.D., Tobin, F.L., Souchet, M. and Brill, A. Direct Submission

TITLE Submitted (02-JUN-1999) Cardiovascular Pharmacology, SmithKline Beecham, 4 rue du Chesnay-Beauregard, Saint-Gregoire 35760, France

FEATURES

Location/Qualifiers

source 1..2152

organism="Homo sapiens"

molecule="mRNA"

db_xref="taxon:9606"

chromosome="1"

map="1p36.23-p34.3 between markers DIS434 and DIS507"

5'..587

/note="25 kDa heat stress protein; cHSP; preferentially expressed in cardiovascular and insulin-sensitive tissues; interacts with C-terminal domain of alpha-filamin or actin-binding-280, Abp-280"

/codon_start=1

/product="cardiovascular heat shock protein"

/protein_id="AA020022.1"

/db_xref="GI:6635998"

/translation="MSHRTSSSTPFRARSRFSSSSSSSTSSASRALPADPMEKALSMFDGSGFMRPSSEPLAPARGAGNKTIGDAVEFAVDYRDSPEDIIVTGNHIEVRAEKLAADGTVMTPAHKQQLPEVDVPTVTSALREDSLTIRRRHPTHEVQOTFETETKI"

ORIGIN

polyA_signal 2115..2120

Query Match 82.7%; Score 1642.8; DB 9; Length 2152;

Best Local Similarity 95.6%; Pred. No. 0;

Matches 1811; Conservative 0; Mismatches 59; Indels 24; Gaps 11;

Qy 33 GCCCATCCCAACCAAGCCCGGTGGGCAAGCAATCAAGACCTGAGAGAGCCCTATGA 92

Db 263 GGCCTTCCCAAGCCCGGTGGGCAAGCAATCAAGACCTGAGAGAGCCCTATGA 322

Qy 93 GTTGGGAG 152

Db 323 GTTGGGAG 379

Qy 153 AACCAACATGAG 212

Db 380 CAACCAATGAG 437

Qy 213 CACCTTACGAGCTACACAG 272

Db 438 TTGGCT-----CAG 486

Qy 273 CTTGGGCTACTGTCGGAG 332

Db 487 CTTGGGCTACTGTCGGAG 543

Qy 333 CAGAACAGCTCCAG 392

Db 544 CAGAACAGCTCCAG 603

Qy 393 TTTTCCTGTCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 452

Db 604 TTTTCCTGTCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 663

Qy 453 TCCAG 512

Db 664 TCCAG 722

Qy 513 CTCCTCAAG 572

Db 723 CTCCTCAAG 782

Qy 573 ATGATGCCAGATTTTGAATTTGGCTCCCTGCTCACTTAATCCAGAGAGAGAGAGAG 632

Db 783 ATGATGCCAGATTTTGAATTTGGCTCCCTGCTCACTTAATCCAGAGAGAGAGAGAG 842

Qy 633 CAGGGAAG 692

Db 843 CAGGGAAG 902

Qy 693 TGGGCAAG 752

Db 903 TGGGCAAG 962

Qy 753 GCGAAG 812

Db 963 GCGAAG 1022

Qy 813 AGGTCAAG 872

Db 1023 AGGTCAAG 1082

Qy 873 AAG 932

Db 1083 AAG 1142

Qy 933 GAGCAG 992

Db 1143 GAGCAG 1202

Qy 993 GCGACATCTGTATATAG 1052

Db 1203 GCGACATCTGTATATAG 1262

Qy 1053 AACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1112

Db 1263 AACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1322

Qy 1113 GGGCTGTCT 1172

Db 1323 GGGCTGTCT 1382

Qy 1173 ATCACTTAACAG 1232

Db 1383 ATCACTTAACAG 1442

Qy 1233 CCGAGCTTTTGGCACTTCCCTCC-GGAGTCAAGCTCTCCATCCATCCCTCTTTAA 1291

Db 1443 CCGAGCTTTTGGCACTTCCCTCC-GGAGTCAAGCTCTCCATCCATCCCTCTTTAA 1502

Db 843 CAGGAGGAGAGATCTAAAGAAACCACTGTGGGTGAGGGAAATGGGACCAGCAGACATA 902
 Qy 693 TGGGCAAGCTCTGCAGAGACAGACAGACAAACCTCTGTATCTATGAAAGTCTGTGAG 752
 Db 903 TGGGCAAGCTCTGCAGAGACAGACAGACAAACCTCTGTATCTATGAAAGTCTGTGAG 962
 Qy 753 GGCAGAGGAGACAGGAGACCTGGAAACCTTTGGCCAGGGGAGTGGAGAGACAGAGGA 812
 Db 963 GGCAGAGGAGACAGGAGACCTGGAAACCTTTGGCCAGGGGAGTGGAGAGACAGAGGA 1022
 Qy 813 AGGTCAAGGAGGAGGAGTGTCTATCTAAGTGAATATTTGCCGAGGGCTCAGCAAGGCC 872
 Db 1023 AGGTCAAGGAGGAGGAGTGTCTATCTAAGTGAATATTTGCCGAGGGCTCAGCAAGGCC 1082
 Qy 873 AAGAGAGAGACAGCGGTGAGCGTAACTTCCCTCTACAGGCTCCAGAGCCCAAGCCAGC 932
 Db 1083 AAGAGAGAGACAGCGGTGAGCGTAACTTCCCTCTACAGGCTCCAGAGCCCAAGCCAGC 1142
 Qy 933 GAGCAGGAGTGGCCGACCAACCCGCTGCCCCAGCCAGCTGGCTGTGCAAGGAGAGCCAT 992
 Db 1143 GAGCAGGAGTGGCCGACCAACCCGCTGCCCCAGCCAGCTGGCTGTGCAAGGAGAGCCAT 1202
 Qy 993 GCCACATCTGTATATATGATGGGGTTTTTCCAAATACAGCTGTGTGTAAGAACTGATGA 1052
 Db 1203 GCCACATCTGTATATATGATGGGGTTTTTCCAAATACAGCTGTGTGTAAGAACTGATGA 1262
 Qy 1053 AACTCTGCGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTG 1112
 Db 1263 AACTCTGCGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTG 1322
 Qy 1113 GGGCTGTGCTTCTCCCTCCACAGGCTGTGTTCTGTTGTTGTTGTTGTTGTTGTTGTTG 1172
 Db 1323 GGGCTGTGCTTCTCCCTCCACAGGCTGTGTTCTGTTGTTGTTGTTGTTGTTGTTGTTG 1382
 Qy 1173 ATCACTTAACAGAGATGGAAGCCAGGGCATTGGATGGGGCTTTGGGTCTCTGAGTTGAG 1232
 Db 1383 ATCACTTAACAGAGATGGAAGCCAGGGCATTGGATGGGGCTTTGGGTCTCTGAGTTGAG 1442
 Qy 1233 CCAGCTTCTGCGACCTTCCCTCC- GGCACTGAGCTGCTCATCATCCCTCTTAA 1291
 Db 1443 CCAGCTTCTGCGACCTTCCCTCC- GGCACTGAGCTGCTCATCATCCCTCTTAA 1502
 Qy 1292 TCTATGATCTATAGGCTGCTGTGTGTGTAACA- AACAACCTTATGCTTGTCTTCAAT 1350
 Db 1503 TCTATGATCTATAGGCTGCTGTGTGTGTAACAACAACCTTATGCTTGTCTTCAAT 1562
 Qy 1351 ACTCAGCATTAACATGATGTTGAGGCGCAATTGAGAGCTTTCTCAATCAGATTTTCAATC 1410
 Db 1563 ACTCAGCATTAACATGATGTTGAGGCGCAATTGAGAGCTTTCTCAATCAGATTTTCAATC 1622
 Qy 1411 TCCATTTTCTTAACCGGGGAAACATCCCGAGCCACTGATGCTGTGCTTTGTCTGAA 1470
 Db 1623 TCCATTTTCTTAACCGGGGAAACATCCCGAGCCACTGATGCTGTGCTTTGTCTGAA 1682
 Qy 1471 GGTATATCTGAACCCAGGAGTGTCAACGCTGCTCAACTGCCAAGCTTGGGAGACTGA 1530
 Db 1683 GGTATATCTGAACCCAGGAGTGTCAACGCTGCTCAACTGCCAAGCTTGGGAGACTGA 1742
 Qy 1531 GAGATATTTCCCTCAATCTACTAGGCTATGCAACCTTCCAGAGCTTTCCAGC 1590
 Db 1743 GAGATATTTCCCTCAATCTACTAGGCTATGCAACCTTCCAGAGCTTTCCAGC 1801
 Qy 1591 TGGGGGATGGGGGAGTCAATAGAAAAGCCCATCTCCATCTGGGATAGGGAACCTTCC 1650
 Db 1802 TGGGGGATGGGGGAGTCAATAGAAAAGCCCATCTCCATCTGGGATAGGGAACCTTCC 1861
 Qy 1651 ATCAGGCTTAACCTTGGGAAATGCTGCTGCCCGAGAGTACTTGTGTTGCTTCCAGC 1710
 Db 1862 ATCAGGCTTAACCTTGGGAAATGCTGCTGCCCGAGAGTACTTGTGTTGCTTCCAGC 1921
 Qy 1711 ATACAGAACAGGCTGAGAGGAGAGGCTGTCTAGTTAGCAGGCTCCAGGAGCAAG 1770
 Db 1922 ATACAGAACAGGCTGAGAGGAGAGGCTGTCTAGTTAGCAGGCTCCAGGAGCAAG 1981

Qy 1771 TCAAGCTCTCCCTCCATGCTCTCTGTGCTAGTGTGCTTAAAGTGGCTTCTCACTCCA 1830
 Db 1982 TCAAGCTCTCCCTCCATGCTCTCTGTGCTAGTGTGCTTAAAGTGGCTTCTCACTCCA 2041
 Qy 1831 CCACTCTGGG- CCCTTGGGGGAGAGACTGGGGAGGGGGCGTGGGAGAGCCCTGAGCGCTGG 1889
 Db 2042 CCACTCTGGGCCCCCTTGGGGAGAGACTGGGGAGGGGGCGTGGGAGAGCCCTGAGCGCTGG 2101
 Qy 1890 AACCTGTATACACATATAAGGACAGTCTCAGAGA 1923
 Db 2102 AACCTGTATACACATATAAGGACAGTCTCAGAGA 2135

Search completed: October 24, 2004, 13:54:30
 Job time : 8565 secs

Qy	170	ACATACAGAAAGCAGGGGTGAGAGGGAAGGATGGGCTCAGTTAGAGAGGGTCCACAG3CA	176
Db	61	ACATACAGAAAGCAGGGGTGAGAGGGAARGTGGCTCAGTTAGCAGAGGGTCCACAGGCA	120
Qy	1769	AGTACGCGCTCTCCCTCCATGCGCTCTCTGATCAGTGTGCGCTTAAGGTGGCGCTCTCACTCC	1822
Db	121	AGTACGCGCTCTCCCTCCATGCGCTCTCTGATCAGTGTGCGCTTAAGGTGGCGCTCTCACTCC	180
Qy	1829	CACCACTCTTGGG-CCCTTGGGGGAGAGACTGGGGAGAGGGGCGCTGGGAGAGACCTGTACGCT	1887
Db	181	CACCACTCTTGGGCGCCCTTGGGGGAGAGACTGGGAGAGGGGCGCTGGGGAARCCCTGTACGCT	240
Qy	1888	GGAACC 1893	
Db	241	GGAACC 246	

RESULT 2

```

US-09-620-312D-669
Sequence 669, Application US/09620312D
Patent No. 6569662

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Yindong
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Weinman, Tom
APPLICANT: Xue, Aйдong
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhimei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784C1P23
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ. ID NOS: 1105
SOFTWARE: pt_fl_genes Version 1.0
SEQ ID NO 669
LENGTH: 719
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (81)..(576)
US-09-620-312D-669

```

Query Match	8.6%;	Score 171.4;	DB 4;	Length 719;
Best Local Similarity	78.2%;	Pred. No. 9.7e-41;		
Matches 287; Conservative	0;	Mismatches 61;	Indels 19;	Gaps 6

Qy	33	GGCCATCCCAACCAAGCCCCGGTGGGACAGGCAATCATTAAGAACCCTAGAGAGCGCTATGA	92
Dd	254	GGCTTCCCAAGCCCCCGCGTGGGACAGGCAATCATTAAGACCCCTAAGAGCGCTATGA	313
Qy	93	GTTTGGGGTGGGAGGTGAGAGACTTCTCACCCCTGAAGAAACACTCATTTGTCACCACTCCAAACA	152
Dd	314	GTTTGGGGTGGGAGGTGAGAGACTTCTCACCTGAAGAAACTCATTTGTCACCACTCCAAACA	370
Qy	153	AACCAACATCGAGGTGCGGGCTGAGAAAGCTGGCGGCTGAACGAGCACTGTACATGAAA	212

Db	371	CAACCAACATCGAGGTGCGCG--CTGAGAAAGCTGGGGGTGACGGCACTGTATGAAACACC	428
Qy	213	CACCTTAACGACTACACCAAGTGCACAGATCGCGGAGAGAGTGAACCCGAGGGTCCGCTGA	272
Db	429	TTGGCT-----CAAAATGCGAA--CTGCCGAGAGAGTGAATCC---GACGTGGTGA	477
Qy	273	CTTCGGCTTCTGTGCGGAGAGACGACACCTTCACTAGTCGCGGACGCCGTTCACCCGATA	332
Db	478	CTCTGG---CTGCGCGGAGAGCGGCACTCTACTATCTCGGACCGGCGCTAACCCGATA	534
Qy	333	CAGAACGCTCGAGCAGACCTTCCGAGCGGAGATCAAAATCTGAGTGCCTCTCCCTTCCC	352
Db	535	CAGAAACGCTCAGGAGACTTCCGAGCGGAGATCAAAATCTGAGTGCCTCTCTGCTCAG	594
Qy	393	TTTCCT 399	
Db	595	TGTGCTT 601	

RESULT 3

```

US-09-513-999C
; Sequence 20093, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J. B.
; APPLICANT: Duclet, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; Patent No. 6783961
; FILE REFERENCE: 59, US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 20093
; LENGTH: 125
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 120
; OTHER INFORMATION: m=a or c
; US-09-513-999C-20093

```

Query Match	6.0%	Score 119.6;	DB 4;	Length 125;
Best Local Similarity	98.3%	Pred. No. 9e-26;		
Matches 119; Conservative	1;	Mismatches	1;	Indels 0; Gaps 0;

Qy	1398	CAGATTACAACTCCCACTTTTCATTAAACGGGGAAACAATCCCGAGCACTAGTGTGTG	1457
Db	1	CAGATTACAACTCCCACTTTTCATTAAACGGGGAAACAATCCCGAGCACTAGTGTGTG	60
Qy	1458	CTTTGTGCTAGGAAGTTAGATCTGAAACCCAGGGGTGTAAACNGCTGCTGAACTCCCCAC	1517
Db	61	CTTTGTGCTAGGAAGTTAGATCTGAAACCCAGGGGTGTAAACNGCTGCTGAACTCCCCAC	120

QY	1518	C	1511
Db	121	C	121

RESULT 4

RESULT 4
US-09-513-999C-2322
Sequence 2322, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclerc, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins

```

? Patent No.6783961
? FILE REFERENCE: 59, US2, REG
? CURRENT APPLICATION NUMBER: US/09/513,999C
? CURRENT FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/122,487
? PRIOR FILING DATE: 1999-02-26
? NUMBER OF SEQ ID NOS: 36681
? SOFTWARE: Patent.pm
? SEQ ID NO 2342
?
? LENGTH: 363
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 75..383
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 131
?
? OTHER INFORMATION: y=c or t
?
? US-09-513-999C-2342

```

	Query Match	Similarity	Score	DB	length
Best Local	108	92.3%	1.9e-20		383
Matches	108	Conservative	0	Mismatches	9
				Indels	0
				Gaps	0

```

RESULT 5
US-09-621-976-2084
Sequence 2084, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: JoBERT, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2084
LENGTH: 463
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: CDS
LOCATION: 118..363
NAME/KEY: misc_feature
LOCATION: 60,355,415
OTHER INFORMATION: n=a, g, c or t
US-09-621-976-2084

```

	Query Match	Similarity	Score	DB	Length
Best Local	Similarity	87.2%	Pred. No. 3.3e-17		
Matches	123	Conservative	0	Mismatches 14	Indels 4; Gaps 2
Qy	33	GCCCATCCCAACCAAGCCCGGTGGAGGCAACATCAA-GACCTTAGAGAGCGCTATG	91		
Db	306	GGCCTTCCCAAGCCCGCGCGGTGGGGGACAGCAACATCAAGGACCTTAGAGAGCGCTATG	365		
Qy	92	AGTTTGGGTGACGTGAGAGACTTCTCACTTGAAGACATCATTTGTACCACTCCAAAC	151		
Db	366	AGTTTGGCGGTGACGTGAGAGACTTCTCACTTGAAGACATCATTTGTACCACTCCCA	422		
Qy	152	AAACCAACATCGAGGTGCGGG	172		

Db 423 ACAACCATCGAGTGGG 443

RESULT 6
 US-08-232-463-14/c
 Sequence 14: Application US/08232463
 Patent No. 5670367
 GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEIFLINER, F.
 APPLICANT: FALGNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMM
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: pTZ9c-F15
 US-08-232-463-14

[illegible]

Patent No. 6281344
 GENERAL INFORMATION:
 APPLICANT: Szoctak, Jack W.
 APPLICANT: Roberts, Richard W.
 APPLICANT: Liu, Rihue
 TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
 TITLE OF INVENTION: FUSIONS
 FILE REFERENCE: 00786/350007
 CURRENT APPLICATION NUMBER: US/09/244,796
 CURRENT FILING DATE: 1999-02-05
 EARLIER APPLICATION NUMBER: 60/035,963
 EARLIER FILING DATE: 1997-01-27
 EARLIER APPLICATION NUMBER: 60/064,491
 EARLIER FILING DATE: 1997-11-06
 EARLIER APPLICATION NUMBER: 09/007,005
 EARLIER FILING DATE: 1998-01-14
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 17
 LENGTH: 289
 TYPE: RNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Translation template
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)...(289)
 OTHER INFORMATION: n = A,T,C or G
 US-09-244-796-17

Query Match 2.1%; Score 42.6; DB 3; Length 289;
 Best Local Similarity 3.8%; Pred. No. 0.011;
 Matches 8; Conservative 101; Mismatches 100; Indels 0; Gaps 0;

QY 627 TGGGTCAGGGAAGATCTAAGAACCCAGCTGTGGTCAGGGAATGGGACCCAG 686
 DB 42 URTAPCARARUTGRNRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 101
 QY 687 GACATATGGCCAGCTGTGACAGACAGACAGACAGACAAACCTGTATGAGCTT 746
 DB 102 NRSNR 161
 QY 747 CTGAGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 806
 DB 162 NRSNR 221
 QY 807 GAGGAGGTCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 835
 DB 222 ARGGRURGRGRURRPARARURURURURURURURURURURURURURURUR 250

RESULT 10
 US-09-962A-6/c
 Sequence 6, Application US/09909962A
 Patent No. 6534272
 GENERAL INFORMATION:
 APPLICANT: POLYCHRONAKOS, Constantin
 APPLICANT: VAFIADIS, Petros
 APPLICANT: GRABIS, Rosemarie
 APPLICANT: OUNISSI-BENKALHA, Houria
 TITLE OF INVENTION: DNA ASSAY FOR THE PREDICTION OF
 TITLE OF INVENTION: AUTOIMMUNE DIABETES
 FILE REFERENCE: 1770-213US-1 FC
 CURRENT APPLICATION NUMBER: US/09/909,962A
 CURRENT FILING DATE: 2001-07-23
 NUMBER OF SEQ ID NOS: 7
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6
 LENGTH: 2130
 TYPE: DNA
 ORGANISM: human
 US-09-909-962A-6

Query Match 2.0%; Score 39.8; DB 4; Length 2130;
 Best Local Similarity 48.2%; Pred. No. 0.22;
 Matches 145; Conservative 0; Mismatches 152; Indels 4; Gaps 1;

QY 257 CCCGAGCGTGGGTGACCTCGGCTACTGTGGGAGAGGAGGAGGAGGAGGAGGAGG 316
 DB 2130 CCCGAGCGCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2071
 QY 317 AGCGGTACCCCGATACAGAAACAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAG 376
 DB 2070 CCCGAGCGCTGTCTGCAACCCCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAG 2011
 QY 377 GTGCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 436
 DB 2010 ACACCCCTGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1951
 QY 437 ACCCATTCAGAGG---TCAGGACATCTCAGCCAGGTTCTAGCCCCCAGGAGCC 492
 DB 1950 CCCCTGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1891
 QY 493 CAGACCCGAGTGAACATCTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 552
 DB 1890 CTGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1831
 QY 553 G 553
 DB 1830 G 1830

RESULT 11
 US-09-962A-7/c
 Sequence 7, Application US/09909962A
 Patent No. 6534272
 GENERAL INFORMATION:
 APPLICANT: POLYCHRONAKOS, Constantin
 APPLICANT: VAFIADIS, Petros
 APPLICANT: GRABIS, Rosemarie
 APPLICANT: OUNISSI-BENKALHA, Houria
 TITLE OF INVENTION: DNA ASSAY FOR THE PREDICTION OF
 TITLE OF INVENTION: AUTOIMMUNE DIABETES
 FILE REFERENCE: 1770-213US-1 FC
 CURRENT APPLICATION NUMBER: US/09/909,962A
 CURRENT FILING DATE: 2001-07-23
 NUMBER OF SEQ ID NOS: 7
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 7
 LENGTH: 2130
 TYPE: DNA
 ORGANISM: human
 US-09-909-962A-7

Query Match 2.0%; Score 39.8; DB 4; Length 2130;
 Best Local Similarity 48.2%; Pred. No. 0.22;
 Matches 145; Conservative 0; Mismatches 152; Indels 4; Gaps 1;

QY 257 CCCGAGCGTGGGTGACCTCGGCTACTGTGGGAGAGGAGGAGGAGGAGGAGGAGG 316
 DB 2130 CCCGAGCGCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2071
 QY 317 AGCGGTACCCCGATACAGAAACAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAG 376
 DB 2070 CCCGAGCGCTGTCTGCAACCCCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAG 2011
 QY 377 GTGCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 436
 DB 2010 ACACCCCTGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1951
 QY 437 ACCCATTCAGAGG---TCAGGACATCTCAGCCAGGTTCTAGCCCCCAGGAGCC 492
 DB 1950 CCCCTGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1891
 QY 493 CAGACCCGAGTGAACATCTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 552

Mon Oct 25 11:39:41 2004

us-09-977-418-7.rn1

Page 7

	Matches	90;	Conservative	0;	Mismatches	86;	Indels	0;	Gaps	0;
QY	366	TCAAAATCTGAGTACCTCTCTCCCTTCCCTTCCCTGTCGCCGCCAGCCGCTGCGCAGCA	425							
Db	311	TCAAAATCTGAGTACCTCTCTCCCTTCCCTTCCCTGTCGCCGCCAGCCGCTGCGCAGCA	252							
QY	426	AAGCCTGCTAACCCCATTAACAAGCTCCAGGACATCTCAGCCCAAGTTCTAGCCCCA	485							
Db	251	CC	192							
QY	486	CGACCCCCAGACCCCGAGTGGACCATCTTCCAAACTAGGGCCCTCCACTCTATCC	541							
Db	191	CC	136							

Search completed: October 24, 2004, 15:44:03
Job time : 179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 23, 2004, 21:13:22 ; Search time 979 Seconds

(without alignments)
10654.341 Million cell updates/sec

Title: US-09-977-418-7

Perfect score: 1987

Sequence: 1 tgaatgaagsgctgtccatg.....cnnncttacttgagagaaa 1987

Scoring table: IDENTITY NJC

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Genesegq_235604:*
2: genesegq1980s:*
3: genesegq1990s:*
4: genesegq2000s:*
5: genesegq2001as:*
6: genesegq2001bs:*
7: genesegq2002as:*
8: genesegq2002bs:*
9: genesegq2003as:*
10: genesegq2003bs:*
11: genesegq2003ds:*
12: genesegq2004s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	99.6	1987	AAf27852	Human NOV
2	1979	99.6	1987	ACD40257	Human bre
3	1979	99.6	1987	ADf66734	CDNA enco
4	1979	99.6	1987	ADf66734	Human NOV
5	1979	99.6	1987	ADf66734	Human NOV
6	1979	99.6	1987	ADf66734	Human NOV
7	1979	99.6	1987	ADf66734	Human NOV
8	1979	99.6	1987	ADf66734	Human NOV
9	1979	99.6	1987	ADf66734	Human NOV
10	1979	99.6	1987	ADf66734	Human NOV
11	1979	99.6	1987	ADf66734	Human NOV
12	1979	99.6	1987	ADf66734	Human NOV
13	1979	99.6	1987	ADf66734	Human NOV
14	1979	99.6	1987	ADf66734	Human NOV
15	1979	99.6	1987	ADf66734	Human NOV
16	1979	99.6	1987	ADf66734	Human NOV
17	1979	99.6	1987	ADf66734	Human NOV
18	1979	99.6	1987	ADf66734	Human NOV
19	1979	99.6	1987	ADf66734	Human NOV
20	1979	99.6	1987	ADf66734	Human NOV
21	1979	99.6	1987	ADf66734	Human NOV

22	414.2	20.8	421	9	ACH47257	ACH47257 Human inf
23	410.4	20.7	412	9	ACH18011	ACH18011 Human adu
24	409.4	20.6	412	9	ACH17535	ACH17535 Human adu
25	402	20.2	402	6	ABL62195	ABL62195 Colon ade
26	402	20.2	402	6	ABK64128	ABK64128 Human ben
27	399.4	20.1	486	9	ACH15596	ACH15596 Human adu
28	384.2	19.3	757	12	ADL30250	ADL30250 3' end of
29	384.2	19.3	757	12	ADL30250	ADL30250 3' end of
30	375	18.9	400	9	ACH16935	ACH16935 Human adu
31	375	18.9	458	4	AA112114	AA112114 Probe #20
32	375	18.9	458	4	AA112114	AA112114 Probe #20
33	375	18.9	458	4	AA112114	AA112114 Probe #20
34	375	18.9	458	4	AA112114	AA112114 Probe #20
35	375	18.9	458	4	AA112114	AA112114 Probe #20
36	375	18.9	458	4	AA112114	AA112114 Probe #20
37	375	18.9	458	4	AA112114	AA112114 Probe #20
38	375	18.9	458	4	AA112114	AA112114 Probe #20
39	375	18.9	458	4	AA112114	AA112114 Probe #20
40	375	18.9	458	4	AA112114	AA112114 Probe #20
41	375	18.9	458	4	AA112114	AA112114 Probe #20
42	375	18.9	458	4	AA112114	AA112114 Probe #20
43	375	18.9	458	4	AA112114	AA112114 Probe #20
44	375	18.9	458	4	AA112114	AA112114 Probe #20
45	375	18.9	458	4	AA112114	AA112114 Probe #20

ALIGNMENTS

RESULT 1	AAf27852	standard; CDNA; 1987 BP.
ID	AAf27852	standard; CDNA; 1987 BP.
AC	AAf27852	
XX	AAf27852	
DT	30-MAR-2001	(first entry)
XX	30-MAR-2001	(first entry)
DE	Human NOV4 CDNA.	
XX	Human NOV4 CDNA.	
XX	Human; NOVX; antiinflammatory; cyrostatic; neuroprotective;	
KW	cerebroprotective; immunomodulator; vulnerary; vasculoprotective;	
KW	hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;	
KW	diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease; ss.	
OS	Homo sapiens.	
XX	Homo sapiens.	
XX	WO200075321-A2.	
XX	14-DEC-2000.	
XX	01-JUN-2000; 2000WO-US015303.	
XX	03-UTN-1999; 99US-0137322P.	
PR	16-MAR-2000; 2000US-0189810P.	
PR	22-MAR-2000; 2000US-0191158P.	
PR	30-MAR-2000; 2000US-0193086P.	
PR	31-MAR-2000; 2000US-00137322.	
XX	(CURA-) CURAGEN CORP.	
XX	Shinketsu RA, Fernandes E, Herrman J, Vernet C;	
PI	WPI; 2001-102403/11.	
DR	P-PSDB; AAB61132.	
XX	New NOVX polypeptides and polynucleotides, useful in gene therapy, as a	
PT	diagnostic marker, protein therapeutic, antibody or small molecule drug	
PT	target for treating immune, proliferative and metabolic diseases and	
XX	wound healing.	
XX	Claim 8; Page 26-27; 194pp; English.	
XX	The present sequence encodes a new isolated polypeptide (NOVX). The NOVX	

CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC treating or preventing NOVX-associated disorders. They are also useful
CC for determining the presence of or a predisposition to a disease
CC associated with altered levels of the NOVX polypeptide or nucleic acid.
CC These NOVX-associated disorders include hyperplasias, tumours,
CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC are especially useful in gene therapy. Specifically, NOVX is useful as a
CC diagnostic marker or prognostic marker, protein therapeutic and antibody
CC target or small molecule drug target to treat disorders in the immune
CC response pathway, thyroid and metabolic diseases, bone metabolic
CC disorders, diseases of the pancreas (e.g. diabetes or digestive
CC disorders), proliferative diseases, or tissue regeneration and
CC development (e.g. wound healing or treatment of burns)

XX Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

Query Match 99.6%; Score 1979; DB 4; Length 1987;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 TGAATGAGGCGCTGTCATGCTCGTCAAGAGCCCATCCAAAGCCCGGCTGGGCA 60
QY 1 TGAATGAGGCGCTGTCATGCTCGTCAAGAGCCCATCCAAAGCCCGGCTGGGCA 60
DB 1 TGAATGAGGCGCTGTCATGCTCGTCAAGAGCCCATCCAAAGCCCGGCTGGGCA 60
QY 61 GGGCAATCAAGACCCCTAGAGAGCGCTATGAGTTGGCGGAGAGAGACTTCTCA 120
DB 61 GGGCAATCAAGACCCCTAGAGAGCGCTATGAGTTGGCGGAGAGAGACTTCTCA 120
QY 121 CCTGAAGACATCATTTGTCAACACCTCCAAACCAACCAACATGAGGTGGGGCTGAGA 180
DB 121 CCTGAAGACATCATTTGTCAACACCTCCAAACCAACCAACATGAGGTGGGGCTGAGA 180
QY 181 AGGTGGCGGCTGAACGGAGACCTGTACATATAAACACTTACGACTACAGAGTGGCACT 240
DB 181 AGGTGGCGGCTGAACGGAGACCTGTACATATAAACACTTACGACTACAGAGTGGCACT 240
QY 241 GCGCGAGAGCTGAGACCCCGAGCGTGGCGGTGACCTCGGCTACTGTGGGAGAGAGCGAG 300
DB 241 GCGCGAGAGCTGAGACCCCGAGCGTGGCGGTGACCTCGGCTACTGTGGGAGAGAGCGAG 300
QY 301 CCTCACTAGTCCGGGACCGCTACCGGATACAGAACAGCTCCAGCAGACCTTCCGGAC 360
DB 301 CCTCACTAGTCCGGGACCGCTACCGGATACAGAACAGCTCCAGCAGACCTTCCGGAC 360
QY 361 GGAAGATCAAAATCTGAGTGCCTCTGCTTCCCTTCCCTGTCGCCGCGCCCAAGGCTGC 420
DB 361 GGAAGATCAAAATCTGAGTGCCTCTGCTTCCCTTCCCTGTCGCCGCGCCCAAGGCTGC 420
QY 421 GAGCAAAAGCTGCTAACCCCATTAACAAGCTCCAGAGCAATCTCAGCCAGGTTCTAGC 480
DB 421 GAGCAAAAGCTGCTAACCCCATTAACAAGCTCCAGAGCAATCTCAGCCAGGTTCTAGC 480
QY 481 CCCCAAGCAACCCCAAGCCAGAGTGGAGCAATCTCCCAAATAGAGGCTCTCACTATTC 540
DB 481 CCCCAAGCAACCCCAAGCCAGAGTGGAGCAATCTCCCAAATAGAGGCTCTCACTATTC 540
QY 541 CAGGGGAGGCTCAGGAGCTCCCTGAGCTGACATGATGCCAGATTTCAGATTGGGCTTC 600
DB 541 CAGGGGAGGCTCAGGAGCTCCCTGAGCTGACATGATGCCAGATTTCAGATTGGGCTTC 600
QY 601 CCTCACTTATTCAGAGTACAGGGGCTGGGGTCAAGGAGAGAGATCTTAAGAACCACT 660
DB 601 CCTCACTTATTCAGAGTACAGGGGCTGGGGTCAAGGAGAGAGATCTTAAGAACCACT 660
QY 661 GTGGGTCAAGGGATGGGACAGAGCATATGGGCAAGCTCTGACAGACAGACAGCA 720
DB 661 GTGGGTCAAGGGATGGGACAGAGCATATGGGCAAGCTCTGACAGACAGACAGCA 720
QY 721 GACAAACCTCTGATCTATGAAGTCTTGACAGGGCAAGGGAGCAAGGAGCTTGAACTT 780
DB 721 GACAAACCTCTGATCTATGAAGTCTTGACAGGGCAAGGGAGCAAGGAGCTTGAACTT 780

QY 781 CTTGGCCAAAGGGAGTGGAGAGACAGAGGAGAGTCAACAGGAGGCTCTATCTAAG 840
DB 781 CTTGGCCAAAGGGAGTGGAGAGACAGAGGAGAGTCAACAGGAGGCTCTATCTAAG 840
QY 841 TGAAGTAAATTTGCCCGAGGGCTCAGCAAGCCCAAGAGAGACAGCTGACGTTAACTT 900
DB 841 TGAAGTAAATTTGCCCGAGGGCTCAGCAAGCCCAAGAGAGACAGCTGACGTTAACTT 900
QY 901 CCCCTTACAGAGCTTCCAAAGCCCAAGCGACAGAGAGGCTGCTGCCACCCCGAGGCC 960
DB 901 CCCCTTACAGAGCTTCCAAAGCCCAAGCGACAGAGAGGCTGCTGCCACCCCGAGGCC 960
QY 961 CCAGCCAGTGGCTGTGTCAGGAGGAGCCATCCATCTGTATATGATGGGGTTTTT 1020
DB 961 CCAGCCAGTGGCTGTGTCAGGAGGAGCCATCCATCTGTATATGATGGGGTTTTT 1020
QY 1021 CCAATACAGCTGGTGTGTAAGAAATGCAATGAAATCCGCTCCGCTGGGCTGAGGG 1080
DB 1021 CCAATACAGCTGGTGTGTAAGAAATGCAATGAAATCCGCTCCGCTGGGCTGAGGG 1080
QY 1081 CCTTCAGGCAAGGCAAGTGGGGTGGGGGTGGGCTGCTTCTCCCTCCACAGGCC 1140
DB 1081 CCTTCAGGCAAGGCAAGTGGGGTGGGGGTGGGCTGCTTCTCCCTCCACAGGCC 1140
QY 1141 TGTGTTCTNTGGGCTGTGCTCCATGACAGAGATTCACCTAAGAGATGGAAGCCAGGCC 1200
DB 1141 TGTGTTCTNTGGGCTGTGCTCCATGACAGAGATTCACCTAAGAGATGGAAGCCAGGCC 1200
QY 1201 ATGATGAGGCTTTGGGCTCTCAGAGTGGAGCCCAAGCTTCTTCCACCTTCCCTCCG 1260
DB 1201 ATGATGAGGCTTTGGGCTCTCAGAGTGGAGCCCAAGCTTCTTCCACCTTCCCTCCG 1260
QY 1261 CAGTCAAGCTTCCATGCAATCCCTCTTTATATGAAATCTATAGAGCTGGTGTGTGA 1320
DB 1261 CAGTCAAGCTTCCATGCAATCCCTCTTTATATGAAATCTATAGAGCTGGTGTGTGA 1320
QY 1321 ACAACACACCCCTATGTTGCTCTTCAATAACTCAGACTTACCATGTTGAGGCCAAT 1380
DB 1321 ACAACACACCCCTATGTTGCTCTTCAATAACTCAGACTTACCATGTTGAGGCCAAT 1380
QY 1381 TCAAGAGCTTTCAAAATCAGATTACAAATCTCCATTTTCAATTAACGGGGAACATCCCG 1440
DB 1381 TCAAGAGCTTTCAAAATCAGATTACAAATCTCCATTTTCAATTAACGGGGAACATCCCG 1440
QY 1441 AGCCAGTGAAGTGTGCTTGTGCACTGAAGGTTGATGTGAACCCAGGAGTGTCAACGCG 1500
DB 1441 AGCCAGTGAAGTGTGCTTGTGCACTGAAGGTTGATGTGAACCCAGGAGTGTCAACGCG 1500
QY 1501 TGCCTCAACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTCACTCTTCA 1560
DB 1501 TGCCTCAACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTCACTCTTCA 1560
QY 1561 GGCTATGACCCCTCCCAAGCTTCCAGCTGGGAGATGGGGAGATCAAGGAAAGGCC 1620
DB 1561 GGCTATGACCCCTCCCAAGCTTCCAGCTGGGAGATGGGGAGATCAAGGAAAGGCC 1620
QY 1621 CCCATCTCCATCTGGAGATGAGGACCTTCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680
DB 1621 CCCATCTCCATCTGGAGATGAGGACCTTCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680
QY 1681 CCCCCAGTACTTGTGTTTGTCTTCCCATATACAGAGAGGAGGAGAGAGGTTGG 1740
DB 1681 CCCCCAGTACTTGTGTTTGTCTTCCCATATACAGAGAGGAGGAGAGAGGTTGG 1740
QY 1741 GTCTCAGTTAGAGGGGCTCCCAAGGCAAGTCAAGCTCTCCCTCATGCTCTGTGTC 1800
DB 1741 GTCTCAGTTAGAGGGGCTCCCAAGGCAAGTCAAGCTCTCCCTCATGCTCTGTGTC 1800
QY 1801 AGTGTCCCTTAAAGGAGCTCTTCACTCCCACTCTGAGCCCTTGGGGAGAGACTGAGG 1860
DB 1801 AGTGTCCCTTAAAGGAGCTCTTCACTCCCACTCTGAGCCCTTGGGGAGAGACTGAGG 1860

QY	1861	AGGGGGCGCTGGAGAGAGCCCTGAGCTGAACTGTATACAACTAAAGGACATCTCAC	1920
Db	1861	AGGGGGCGCTGGAGAGAGCCCTGAGCTGAACTGTATACAACTAAAGGACATCTCAC	1920
QY	1921	AGACNTCTGGAGAGCGCGCTGCGCGAGAGTTCCTCAACTTAAAGGCGAGCGGCGNNNNCTTACTTG	1960
Db	1921	AGACNTCTGGAGAGCGCGCTGCGCGAGAGTTCCTCAACTTAAAGGCGAGCGGCGNNNNCTTACTTG	1960
QY	1981	AGAGAAA	1987
Db	1981	AGAGAAA	1987
DE	RESULT 2		
XX	ACD40257	standard; DNA; 1987 BP.	
XX	ACD40257;		
XX	ACD40257;		
XX	03-SEP-2003	(first entry)	
DE	Human breast tumour associated protein 47-like polypeptide NOV4 DNA.		
KM	Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;		
KM	pancreatic cancer; uterine cancer; organ transplantation disorder; ds;		
KM	cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;		
KM	ischemic heart disease; haemorrhage; peripheral vascular disease; gene;		
KM	thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;		
KM	tissue regeneration; wound healing; hyperproliferative disorder;		
KM	periorbitis; neural disorder; Parkinson's disease; Alzheimer's disease;		
KM	Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;		
KM	AUS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;		
KM	neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;		
KM	gene therapy; epilepsy; breast tumour associated protein 47.		
OS	Homo sapiens.		
PN	US2003027158-A1.		
XX	06-FEB-2003.		
XX	15-OCT-2001; 2001US-00977418.		
XX	03-JUN-1999; 98US-0137322P.		
PR	16-MAR-2000; 2000US-0189810P.		
PR	22-MAR-2000; 2000US-0191158P.		
PR	30-MAR-2000; 2000US-01931086P.		
PR	03-MAY-2000; 2000US-0201388P.		
PR	31-MAY-2000; 2000US-00564411.		
XX	(CURA-) CURAGEN CORP.		
XX	Shinketsu RA, Fernandes E, Herrman J, Vernet C;		
XX	WPI; 2003-492028/46.		
DR	P-FSDB; ABO23236.		
XX	New nucleic acid sequence encoding a human breast tumor-associated		
PT	protein 47-like polypeptide, useful for treating cardiovascular		
PT	disorders, neural disorders, diabetes mellitus and cancers.		
XX	Claim 8; Page 16-17; 100pp; English.		
CC	The invention relates to a new isolated NOV4 nucleic acid. The nucleic		
CC	acid is useful for identifying a compound that binds the nucleic acid.		
CC	The nucleic acid is useful in gene therapy, in screening assays, in		
CC	detection assays e.g. chromosomal mapping, cell and tissue typing and		
CC	forensic biology, predictive medicine e.g. diagnostic assays, prognostic		
CC	assays, monitoring clinical trials and pharmacogenomics and methods of		
CC	treatment including therapeutic and prophylactic. The nucleic acid is		
CC	also useful for expressing NOV4 protein. The nucleic acid is also useful		
CC	to provide polynucleotide reagents e.g. labeled probes that are useful		
CC	in an in situ hybridisation technique, for identifying a specific tissue		

Query Match	Best Local Similarity	Score 1997;	DB 9;	Length 1987;
Matches 1987;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0
Qy	1	TGATTAAGAGCGCTGTCATAGCTCGTCAAAAGCCATCCCAACCAAGCCCGGTGAGCA	60	
Db	1	TGATTAAGAGCGCTGTCATAGCTCGTCAAAAGCCATCCCAACCAAGCCCGGTGAGCA	60	
Qy	61	GCGACATCAAGACCCCTTAGAGAGACCGCTTAGATTGGCGTGAAGCTGAGACATTCTCA	120	
Db	61	GCGACATCAAGACCCCTTAGAGAGACCGCTTAGATTGGCGTGAAGCTGAGACATTCTCA	120	
Qy	61	GCGACATCAAGACCCCTTAGAGAGACCGCTTAGATTGGCGTGAAGCTGAGACATTCTCA	120	
Db	61	GCGACATCAAGACCCCTTAGAGAGACCGCTTAGATTGGCGTGAAGCTGAGACATTCTCA	120	
Qy	121	CCGGAAGACATCATTTGTACACACCTCCGAAACAAACCAACATCGAGGTGCGGGGTGAGAA	180	
Db	121	CCGGAAGACATCATTTGTACACACCTCCGAAACAAACCAACATCGAGGTGCGGGGTGAGAA	180	
Qy	181	AGCTGCGGCTGACGAGACACTGTACATGAACAACCTTAGACTTACCAAGTGCAGACT	240	
Db	181	AGCTGCGGCTGACGAGACACTGTACATGAACAACCTTAGACTTACCAAGTGCAGACT	240	
Qy	241	GCGGAGAGACGTGACACCCGAGCGTGGCTGACCTCGGCTACTGTGGGAGAGACGAGCAG	300	
Db	241	GCGGAGAGACGTGACACCCGAGCGTGGCTGACCTCGGCTACTGTGGGAGAGACGAGCAG	300	
Qy	301	CGTCACTAGTGGGAGACGCGTCAACCCGCTATACAGAAACGTCGACAGAGACCTTCGGAC	360	
Db	301	CGTCACTAGTGGGAGACGCGTCAACCCGCTATACAGAAACGTCGACAGAGACCTTCGGAC	360	
Qy	361	GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCGCCCGCCACGCTGC	420	
Db	361	GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCGCCCGCCACGCTGC	420	
Qy	421	CACGAAAGGCTGGCTAACCCCATTTACAAACAGTCCAGGACACTTCAGCCGAGTTTCAGC	480	
Db	421	CACGAAAGGCTGGCTAACCCCATTTACAAACAGTCCAGGACACTTCAGCCGAGTTTCAGC	480	
Qy	481	CCCCAGACACCCACAGACCCAGGTGAGACATCTTCCCAACTAGGGCTCTTCACTTATC	540	
Db	481	CCCCAGACACCCACAGACCCAGGTGAGACATCTTCCCAACTAGGGCTCTTCACTTATC	540	
Qy	541	CAGGAGAGGCTCAGGACTCTCCCTGCTGACACAGAGGCCAGATTTCAGATTGGGCTC	600	
Db	541	CAGGAGAGGCTCAGGACTCTCCCTGCTGACACAGAGGCCAGATTTCAGATTGGGCTC	600	
Qy	601	CGTCACTTATCCAGAGTACAGGGCTGGGGTCAAGGAGAGGAAGATCTAAAGAACCACT	660	
Db	601	CGTCACTTATCCAGAGTACAGGGCTGGGGTCAAGGAGAGGAAGATCTAAAGAACCACT	660	
Qy	661	GTGGGTGAGGGAGATTGGAGACAGACAGACATATGGGACAGCTCTGACAGACAGACAGCA	720	

```
Db 661 GTGGGTGAGGGGAAATGGGACCAAGCAGACATATGGGCAAGCTCTGAGAGACAGACAGACA 720
Qy 721 GACAAACCCCTGATCTATGAAGTCTCTGAGGGGCAAGGGGACCAAGGACCTGGAAACCT 780
Db 721 GAAACACCCCTCTATCTATGAAGTCTCTGAGGGGCAAGGGGACCAAGGACCTGGAAACCT 780
Qy 781 CTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGATGCTTATCTAAG 840
Db 781 CTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGATGCTTATCTAAG 840
Qy 841 TGAACCTAATTTGGCCAGGGCTCAGCAAGGCCAAGGAGAGACAGCCGTGAACGCTAACTT 900
Db 841 TGAACCTAATTTGGCCAGGGCTCAGCAAGGCCAAGGAGAGACAGCCGTGAACGCTAACTT 900
Qy 901 CCCCTTACAGCCTTCCAAAGCCCAAGCAGAGAGAGGCTGCTGCCACCCCGTGGCC 960
Db 901 CCCCTTACAGCCTTCCAAAGCCCAAGCAGAGAGAGGCTGCTGCCACCCCGTGGCC 960
Qy 961 CCAGCCAGCTGGTGTGGCCAGGGCAGAGCAGCATCTGTATATATAGATGGGTTTTT 1020
Db 961 CCAGCCAGCTGGTGTGGCCAGGGCAGAGCAGCATCTGTATATATAGATGGGTTTTT 1020
Qy 1021 CCATACAGCTGGTGTGAAAAAACTGCATGAAACTCTGCGCGTCTGCGCTGCTGGGG 1080
Db 1021 CCATACAGCTGGTGTGAAAAAACTGCATGAAACTCTGCGCGTCTGCGCTGCTGGGG 1080
Qy 1081 CCTCCAGGCAAGGCCAAGTGGGGTGGGGGTGGGCTGCTCTTCTCCCTCCACAGGCC 1140
Db 1081 CCTCCAGGCAAGGCCAAGTGGGGTGGGGGTGGGCTGCTCTTCTCCCTCCACAGGCC 1140
Qy 1141 TGTGTCNTGGGGCTGCTCCCAATGACAGAGATCACTCAACAGAGATGGAAGCCAGGGC 1200
Db 1141 TGTGTCNTGGGGCTGCTCCCAATGACAGAGATCACTCAACAGAGATGGAAGCCAGGGC 1200
Qy 1201 ATGATGAGGAGCTTTGGGTCCTGAGAGTTGAGCCCAAGCTTCTGGCACCTTCCCTCCGG 1260
Db 1201 ATGATGAGGAGCTTTGGGTCCTGAGAGTTGAGCCCAAGCTTCTGGCACCTTCCCTCCGG 1260
Qy 1261 CAGTACGCTCTCATCCATCCCTCTTAACTATGAACTCATAGCTCGCTGCTGCTTA 1320
Db 1261 CAGTACGCTCTCATCCATCCCTCTTAACTATGAACTCATAGCTCGCTGCTGCTTA 1320
Qy 1321 ACAACACACCCCTATGTTGCTTCAAAATACAGCATTAACATTTGTTGAGGCCAAT 1380
Db 1321 ACAACACACCCCTATGTTGCTTCAAAATACAGCATTAACATTTGTTGAGGCCAAT 1380
Qy 1381 TCAGAGCTTTCTCAATCAGATTATACATCTCCATTTTCAATTAAGGGAACATCCCGC 1440
Db 1381 TCAGAGCTTTCTCAATCAGATTATACATCTCCATTTTCAATTAAGGGAACATCCCGC 1440
Qy 1441 AGCAGCTGAGTCTGCTTGTCTCACTGAAGTATGATCTGAACCAAGGCTGTCAACGCG 1500
Db 1441 AGCAGCTGAGTCTGCTTGTCTCACTGAAGTATGATCTGAACCAAGGCTGTCAACGCG 1500
Qy 1501 TGTCTCTCAATCTCCCACTCTGAGGAGTATTTCCCTCATTTCTAAGCTCTTAA 1560
Db 1501 TGTCTCTCAATCTCCCACTCTGAGGAGTATTTCCCTCATTTCTAAGCTCTTAA 1560
Qy 1561 GGGTATGACCCCTCCCAACGCTTCAAGCTGGGGAGTGGGGGGAGTATAGAAAACCC 1620
Db 1561 GGGTATGACCCCTCCCAACGCTTCAAGCTGGGGAGTGGGGGGAGTATAGAAAACCC 1620
Qy 1621 CCCATCTCCCATCTGAGGATAGGAGCCTTCATCAGCCTTAAACCTGGGAAAATGCTGCTG 1680
Db 1621 CCCATCTCCCATCTGAGGATAGGAGCCTTCATCAGCCTTAAACCTGGGAAAATGCTGCTG 1680
Qy 1681 CCCCAGAGACTCTTGTGTTTGTCTCCCAATAGAGAGAGAGGAGGAGGAGGAGGAGG 1740
Db 1681 CCCCAGAGACTCTTGTGTTTGTCTCCCAATAGAGAGAGAGGAGGAGGAGGAGGAGG 1740
Qy 1741 GTCTCAGTTAGAGAGGAGTCCCAAGGAGCAAGTCAAGCTCTCCCTCATAGCTCTTGCTC 1800
Db 1741 GTCTCAGTTAGAGAGGAGTCCCAAGGAGCAAGTCAAGCTCTCCCTCATAGCTCTTGCTC 1800

Db 1741 GTCTCAGTTAGAGAGGAGTCCCAAGGAGCAAGTCAAGCTCTCCCTCATAGCTCTTGCTC 1800
Qy 1801 AGTGTGCTTAAAGGTGAGCTCTCTCACTCCACACTGCTGGGCGCTTGGGGAGAGACTGGGG 1860
Db 1801 AGTGTGCTTAAAGGTGAGCTCTCTCACTCCACACTGCTGGGCGCTTGGGGAGAGACTGGGG 1860
Qy 1861 AGGGGCGCTGGGAGAGACCTCTGACGCTGGAACCTGTATACAAATAAGACAGTCTCAC 1920
Db 1861 AGGGGCGCTGGGAGAGACCTCTGACGCTGGAACCTGTATACAAATAAGACAGTCTCAC 1920
Qy 1921 AGACNTCTGAGAGCGCGCTGCGCNGAGAGTCTCAACTTTAGGGGAGGAGGAGGAGGAGG 1980
Db 1921 AGACNTCTGAGAGCGCGCTGCGCNGAGAGTCTCAACTTTAGGGGAGGAGGAGGAGGAGG 1980
Qy 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 3
ADF6734
ID ADF6734 standard; cDNA; 1987 BP.
XX
XX ADF6734:
AC
XX
XX 12-FEB-2004 (first entry)
DT
XX
XX cDNA encoding novel human protein NOV4.
DE
XX
XX cytosolic; hepatotropic; vulnery; antiproliferic; osteopathic;
XX antihistytic; antihistocytic; haemostatic; vasotropic;
XX thrombolytic; antidiabetic; hypotensive; dermatological;
XX immunosuppressive; antineoplastic; immunostimulant; fungicide;
XX virucide; protozoacide; neuroprotective; antineoplastic; antihistytic;
XX antineoplastic; antiparkinsonian; nootropic; anticonvulsant;
XX NOV4 modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
XX psoriasis; tissue hypertrophy; osteoarthritis;
XX atherosclerotic plaque formation; haemorrhage; ischaemic disease;
XX thrombosis; diabetes mellitus; hypertension; hypothyroidism;
XX immune deficiency; severe combined immunodeficiency; SCID; infection;
XX malaria; candidiasis; autoimmune disorder; connective tissue disease;
XX multiple sclerosis; systemic lupus; erythematous; rheumatoid arthritis;
XX autoimmune pulmonary inflammation; Guillain-Barre syndrome;
XX autoimmune thyroiditis; insulin dependent diabetes mellitus;
XX myasthenia gravis; graft-versus-host disease;
XX autoimmune inflammatory eye disease; asthma; haematopoiesis;
XX tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
XX periodontal disease; Alzheimer's disease; Parkinson's disease;
XX Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
XX human; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX US2003199103-A1.
PN
XX
XX 23-OCT-2003.
PD
XX
XX 15-OCT-2001; 2001US-00977639.
PF
XX
XX 03-JUN-1999; 99US-0137322P.
PR 15-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Fernandes E, Herrman J, Vernet C;
PI
XX
XX WPI; 2004-021196/02.
DR
XX
XX P-PSDB; ADF6735.
XX
```


PT Novel substantially NOX polypeptide useful for diagnosing, preventing
PT and treating diseases e.g., cancer, multiple sclerosis, systemic lupus
erythematosus.

XX
XX
PS Disclosure; SEQ ID NO 7; 165pp; English.

CC The invention describes a substantially purified polypeptide (I) having
CC amino acid sequence chosen from a fully defined NOX sequence (SI) of 708
CC amino acids as given in the specification, or polypeptide having one or
CC more conservative amino acid substitutions of (SI), or mutant or variant
CC of (SI). (I) having (SI) is useful for diagnosing a pathological
CC condition associated with (I) or its activity in a subject e.g. cancer.
CC (I) useful in treatment of cancer, hyperproliferative diseases,
CC cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis,
CC atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal
CC disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I)
CC is useful in treatment of various immune deficiencies and disorders such
CC as severe combined immunodeficiency (SCID), bacterial infection, viral
CC infection such as herpes viral infection, protozoan infection such as
CC malaria, fungal infection such as candidiasis. (I) is also useful in
CC treating autoimmune disorders such as connective tissue disease, multiple
CC sclerosis, systemic lupus, erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune
CC thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis,
CC graft-versus-host disease and autoimmune inflammatory eye disease and
CC asthma. (I) useful in regulation of haematopoiesis, regeneration and
CC tissue growth of bone, cartilage, tendon, ligament and useful for wound
CC healing and tissue repair. (I) is also useful in treatment of burns,
CC incisions and ulcers. (I) also useful in treatment of periodontal
CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective
CC antitumor and anti-inflammatory activity. This sequence encodes a novel
CC human NOX protein.

XX
XX
SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

Query Match 99.6%; Score 1979; DB 12; Length 1987;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAGGCTCTTCATGCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGCA 60
Db 1 TGAATGAGGCTCTTCATGCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGCA 60

QY 61 GGCACATCAAGACCTTAGAGAGAGCTATGAGTTGCGTGGAGCTGAGAGACTTCTCA 120
Db 61 GGCACATCAAGACCTTAGAGAGAGCTATGAGTTGCGTGGAGCTGAGAGACTTCTCA 120

QY 121 CTTGAGAGCATATGTTGACCACTCCAAACCAACCAACATCGAGTGGCGGCTGAGAA 180
Db 121 CTTGAGAGCATATGTTGACCACTCCAAACCAACCAACATCGAGTGGCGGCTGAGAA 180

QY 121 AGCTGGGGCTGAGAGAGACATGATGAAACCTTAGAGCTAGACAGAGTGCAGACT 240
Db 181 AGCTGGGGCTGAGAGAGACATGATGAAACCTTAGAGCTAGACAGAGTGCAGACT 240

QY 241 GCCGAGAGACGTGAGACCCGAGCGTGGCGTACCTCGGCTACTGTGCGGAGAGAGCAG 300
Db 241 GCCGAGAGACGTGAGACCCGAGCGTGGCGTACCTCGGCTACTGTGCGGAGAGAGCAG 300

QY 301 CCTACTAGTCCGGGACGCGTACCCGACATACAGAACAGTCCAGAGACCTTCGGGAC 360
Db 301 CCTACTAGTCCGGGACGCGTACCCGACATACAGAACAGTCCAGAGACCTTCGGGAC 360

QY 361 GGAGATCAAAATCTGAGTGTCTCTCCCTTCCCTTCCCTGCCCCCGCCCAAGCCTGC 420
Db 361 GGAGATCAAAATCTGAGTGTCTCTCCCTTCCCTTCCCTGCCCCCGCCCAAGCCTGC 420

QY 421 CAGCAAGCTCTGCTAACCCCATTAACAAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
Db 421 CAGCAAGCTCTGCTAACCCCATTAACAAGCTCCAGACATCTCAGCCAGGTTCTAGC 480

QY 481 CCCACGACCCCGAGACCCCGAGTGAACATCTCCAAACTAGGCGCTTCACTATC 540

Db 481 CCCACGACCCCGAGACCCCGAGTGAACATCTCCAAACTAGGCGCTTCACTATC 540

QY 541 CAGGACGAGCCAGGAGCTCCCTGAGTGCATGATGATCCAGATTTGATTTGGCTTC 600
Db 541 CAGGACGAGCCAGGAGCTCCCTGAGTGCATGATGATCCAGATTTGATTTGGCTTC 600

QY 601 CCTCACTTAATCCAGATTAACAGGCTGGGCTCAGGAGAGAAATCTTAAGAACCTACT 660
Db 601 CCTCACTTAATCCAGATTAACAGGCTGGGCTCAGGAGAGAAATCTTAAGAACCTACT 660

QY 661 GTGGGTCAAGGAGATGGAGACAGCAGACATATATGGGCAAGCTCTGACAGACAGACA 720
Db 661 GTGGGTCAAGGAGATGGAGACAGCAGACATATATGGGCAAGCTCTGACAGACAGACA 720

QY 721 GACAAACCTCTGATCTATGAAATCTCTGACAGGCAAGGAGACAGGAGCTGAACTCT 780
Db 721 GACAAACCTCTGATCTATGAAATCTCTGACAGGCAAGGAGACAGGAGCTGAACTCT 780

QY 781 CTGGCCAAAGGAGATGGAGAGACAGAGGAGGTCAAGGCAAGGAGGCTGATCTAAG 840
Db 781 CTGGCCAAAGGAGATGGAGAGACAGAGGAGGTCAAGGCAAGGAGGCTGATCTAAG 840

QY 841 TGAATCTAATGCTCCGAGGCTCAGCAAGGCTCAGAGAGACAGGCTGAGGTTAACTT 900
Db 841 TGAATCTAATGCTCCGAGGCTCAGCAAGGCTCAGAGAGACAGGCTGAGGTTAACTT 900

QY 901 CCCCTCTACAGCTCTCAGAGCCCAAGGCTCAGAGAGAGCTGCTGCCACCCCTGCCC 960
Db 901 CCCCTCTACAGCTCTCAGAGCCCAAGGCTCAGAGAGAGCTGCTGCCACCCCTGCCC 960

QY 961 CCAGCAGCTGCTGAGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CCAGCAGCTGCTGAGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

QY 1021 CCAATACAGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1080
Db 1021 CCAATACAGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1080

QY 1081 CCTCAGGCAAGGAG 1140
Db 1081 CCTCAGGCAAGGAG 1140

QY 1141 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
Db 1141 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200

QY 1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTGGCACCTTCCCTCGG 1260
Db 1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTGGCACCTTCCCTCGG 1260

QY 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTATGATCTATGATGATGATGATGAT 1320
Db 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTATGATCTATGATGATGATGATGAT 1320

QY 1321 ACAACACACCCCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380
Db 1321 ACAACACACCCCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380

QY 1381 TCGAGAGCTTTCAATGAGATTTTCAATCTCATTTTCAATGAGAGAGAGAGAGAGAG 1440
Db 1381 TCGAGAGCTTTCAATGAGATTTTCAATCTCATTTTCAATGAGAGAGAGAGAGAGAG 1440

QY 1441 AGCAGTGAAGTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1500
Db 1441 AGCAGTGAAGTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1500

QY 1501 TGTCTCACTCCCACTCTGGGACATGAGAGATTTTCCCTCAATCTACTCTCA 1560
Db 1501 TGTCTCACTCCCACTCTGGGACATGAGAGATTTTCCCTCAATCTACTCTCA 1560

QY 1561 GGTATGACCCCTCCCAAGTCTTTCAGCTGGGAGATGGGAGATGAGAGAGAGAGAG 1620

Dh 1561 GCGTATGACACCCCTCCCACTGCTTCCAGCTGGGGAGTGGGGAGTCAATAGAAAAAGCC 1620
Qy 1621 CCCATCTCCCATCTGGAGATAGAGACCTTCATCAGACCTTAACCTTGGGAAATGCTGCTG 1680
Db 1621 CCCATCTCCCATCTGGAGATAGAGACCTTCATCAGACCTTAACCTTGGGAAATGCTGCTG 1680
Qy 1681 CCCCCAGTACTCTTGTCTTCTCTCCCATCAGAGAGAGGGTGGAGGGAGAGGCTG 1740
Db 1681 CCCCCAGTACTCTTGTCTTCTCTCCCATCAGAGAGAGGGTGGAGGGAGAGGCTG 1740
Qy 1741 GTCTCAGTATGACAGGGGCTCCCAAGGCAAGTACAGCTCTCCCTCCATCAGCTCTGCTGCTC 1800
Db 1741 GTCTCAGTATGACAGGGGCTCCCAAGGCAAGTACAGCTCTCCCTCCATCAGCTCTGCTGCTC 1800
Qy 1801 AGTGTGCTTATGAGGCTGAGCTCTTCTCAGTCCCACTCTGGGCTCTGGGGAGAGTGGGG 1860
Db 1801 AGTGTGCTTATGAGGCTGAGCTCTTCTCAGTCCCACTCTGGGCTCTGGGGAGAGTGGGG 1860
Qy 1861 AGGGGGGCGTGGAGAGACCTGAGCTGAGCTGAGCTGATACAAATTAAGAGAGTCTAC 1920
Db 1861 AGGGGGGCGTGGAGAGACCTGAGCTGAGCTGAGCTGATACAAATTAAGAGAGTCTAC 1920
Qy 1921 AGACNTCTGAGAGGCGCTGCGCNGAGATTCTCAACTTATGAGGAGGCGNNNCTTACTTG 1980
Db 1921 AGACNTCTGAGAGGCGCTGCGCNGAGATTCTCAACTTATGAGGAGGCGNNNCTTACTTG 1980
Qy 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 4
AD060244
ID AD060244 standard; DNA; 1987 BP.
AC AD060244;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human NOV4 DNA.
XX
KW Human; NOVX protein; cancer; hyperproliferative disease; cirrhosis;
KW keloid; psoriasis; tissue hypertrophy; osteoarthritis;
KW atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
KW renal disease; thrombosis; diabetes mellitus; hypertension;
KW hypothyroidism; severe combined immunodeficiency; SCID; infection;
KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KW autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
KW peridontal disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KW haematopoiesis; wound healing; tissue repair; antitumor;
KW antiinflammatory; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 991..1449
FT /tag= b
FT /product= "NOVX protein"
FT /transl_except= (pos:1147..1149, aa:Xaa)
FT /note= "Xaa corresponds to any amino acid"
FT sig_peptide 991..1065
FT /tag= a
FT /tag= 1066..1446
FT /tag= c
FT /product= "Mature NOVX protein"
XX
XX US2003134430-A1.
XX
XX 17-JUL-2003.

XX
PF 15-OCT-2001; 2001US-00977751.
XX
XX 03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0181158P.
PR 30-MAR-2000; 2000US-0191086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shinkets RA, Fernandes E, Herrman J, Vernet C;
XX
XX WPI; 2004-068928/07.
XX
XX P-PSDB; AD060245.
XX
XX Novel substantially purified NOVX polypeptide for treating severe
XX combined immunodeficiency, candidiasis, cancer, asthma, multiple
XX sclerosis, systemic lupus erythematosus.
XX
XX Disclosure; SEQ ID NO 7; 155pp; English.
XX
XX The invention relates to human NOVX polypeptides and polynucleotides.
XX NOVX sequences are useful in the treatment of cancer, hyperproliferative
XX diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
XX osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
XX heart or renal disease, thrombosis, diabetes mellitus, hypertension,
XX hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
XX Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
XX and disorders such as severe combined immunodeficiency (SCID), bacterial
XX infection, viral infection e.g. herpes viral infection, protozoan
XX infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
XX disorders such as connective tissue disease, multiple sclerosis, systemic
XX lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
XX inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
XX dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
XX and autoimmune inflammatory eye disease. The invention is useful in
XX regeneration and tissue growth of bone, cartilage, tendon, ligament,
XX haematopoiesis regulation, wound healing and tissue repair. Sequences of
XX the invention also exhibit antitumor and antiinflammatory activities. The
XX present sequence is human NOVX DNA.
XX
XX Sequence 1987 BP; 448 A; 613 C; 532 G; 366 T; 0 U; 8 Other;
XX
XX Query Match 99.6%; Score 1979; DB 12; Length 1987;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGAATGAAGGCTGTCCATGCTGTCATCAAGAGCCCATCCCAAGCCCGGTGGGA 60
Db 1 TGAATGAAGGCTGTCCATGCTGTCATCAAGAGCCCATCCCAAGCCCGGTGGGA 60
Qy 61 GGCACATCAAGACCTTAGAGAGCGCTTATGAGTTTCCGCTGAGCTGAGAACTTTCTCA 120
Db 61 GGCACATCAAGACCTTAGAGAGCGCTTATGAGTTTCCGCTGAGCTGAGAACTTTCTCA 120
Qy 121 CTTGAAGACATCATTTGTCAACACCTCCAAACCAACCAATGAGGTGGGGCTGAGAA 180
Db 121 CTTGAAGACATCATTTGTCAACACCTCCAAACCAACCAATGAGGTGGGGCTGAGAA 180
Qy 181 AGCTGCGGCTGAGAGCAGACTGTACATGAAACACCTTACGACTACACAAAGTGCAGACT 240
Db 181 AGCTGCGGCTGAGAGCAGACTGTACATGAAACACCTTACGACTACACAAAGTGCAGACT 240
Qy 241 GCGGAGAGACGTGACCCCGAGCGTGGGTGACCTCGCTACTGTCCGAGAGACGAGACG 300
Db 241 GCGGAGAGACGTGACCCCGAGCGTGGGTGACCTCGCTACTGTCCGAGAGACGAGACG 300
Qy 301 CCTCACTAGTCCGGGAGCGCTACCCGAGTACAGAAACAGTCCAGAGACCTTCCGAGC 360
Db 301 CCTCACTAGTCCGGGAGCGCTACCCGAGTACAGAAACAGTCCAGAGACCTTCCGAGC 360

QY 361 GGAGATCAAAATCTGAGTGTCTCCCTCCCTTCCCTGTCGCCCCCGCCAGGCTGC 420
 Db 361 GGAGATCAAAATCTGAGTGTCTCCCTCCCTTCCCTGTCGCCCCCGCCAGGCTGC 420
 QY 421 CAGGAAAGCTCTGCTAACCCCATTAACAAGCTCCAGGACATCTCAGCCAGTTCAGC 480
 Db 421 CAGGAAAGCTCTGCTAACCCCATTAACAAGCTCCAGGACATCTCAGCCAGTTCAGC 480
 QY 481 CCCCACGACCCCGACAGCCCGAGTGGACATCTCTCCAAATTAGGCGCTTCATCTATC 540
 Db 481 CCCCACGACCCCGACAGCCCGAGTGGACATCTCTCCAAATTAGGCGCTTCATCTATC 540
 QY 541 CAGGACGAGGCGAGGAGCTCCCTGGGCTTACACATGATGCCCCAGATTTCAGATTGGGCTC 600
 Db 541 CAGGACGAGGCGAGGAGCTCCCTGGGCTTACACATGATGCCCCAGATTTCAGATTGGGCTC 600
 QY 601 CGTCACTTAATCCAGAGTACAGGAGGCTGGGCTCAGGGAAGAGATCTTAAGAACCCACT 660
 Db 601 CGTCACTTAATCCAGAGTACAGGAGGCTGGGCTCAGGGAAGAGATCTTAAGAACCCACT 660
 QY 661 GTGGGTACGGGGAAATGGGACCAAGACGATATGAGGCAAGCTCTGACAGGACAGACACA 720
 Db 661 GTGGGTACGGGGAAATGGGACCAAGACGATATGAGGCAAGCTCTGACAGGACAGACACA 720
 QY 721 GACAAACCCCTCTGATCTATGAAATCTCTGACAGGCGCAAGGAGCCAGGAGCTGGACCT 780
 Db 721 GACAAACCCCTCTGATCTATGAAATCTCTGACAGGCGCAAGGAGCCAGGAGCTGGACCT 780
 QY 781 CTGGGCGCAAGGAGGAGTGGGAGAGACAGAGGGAAGTCAAGGCAAGGAGTCTATCTAG 840
 Db 781 CTGGGCGCAAGGAGGAGTGGGAGAGACAGAGGGAAGTCAAGGCAAGGAGTCTATCTAG 840
 QY 841 TGAACCTAATGCGCCAGGAGGCTCAGCAAGGCGCAAGGAGGAGGAGGAGGAGTCTAG 900
 Db 841 TGAACCTAATGCGCCAGGAGGCTCAGCAAGGCGCAAGGAGGAGGAGGAGGAGTCTAG 900
 QY 901 CCCCTCTACAGCTCTCCAGAGCCCAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
 Db 901 CCCCTCTACAGCTCTCCAGAGCCCAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
 QY 961 CCAGCGAGGCTGCTGTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
 Db 961 CCAGCGAGGCTGCTGTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
 QY 1021 CCAATACAGCTGTGCTGTGAAAACTGCGATGAAACTCTGCGCTGCTGCTGCTGCTG 1080
 Db 1021 CCAATACAGCTGTGCTGTGAAAACTGCGATGAAACTCTGCGCTGCTGCTGCTGCTG 1080
 QY 1081 CCTCAGGCGAGGCGCAAGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGG 1140
 Db 1081 CCTCAGGCGAGGCGCAAGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGG 1140
 QY 1141 TGTGTTCTGAGGAGGCTGCTCCATGACAGAGTACCTTAACAGAGATGAGGAGGAGG 1200
 Db 1141 TGTGTTCTGAGGAGGCTGCTCCATGACAGAGTACCTTAACAGAGATGAGGAGGAGG 1200
 QY 1201 ATGATGAGGAGGCTGTTGGGCTCTGAGAGTGGAGCCCAAGTCTTTCGCACTTCCCTCGG 1260
 Db 1201 ATGATGAGGAGGCTGTTGGGCTCTGAGAGTGGAGCCCAAGTCTTTCGCACTTCCCTCGG 1260
 QY 1261 CAGTCAGGCTTCATCCATCCCTCCCTTAACTATGATGATGATGATGATGATGATGATG 1320
 Db 1261 CAGTCAGGCTTCATCCATCCCTCCCTTAACTATGATGATGATGATGATGATGATGATG 1320
 QY 1321 ACAACACACCCCTATGCTGTTGCTTCAAAATCTCAGCATTAACATGAGTGGAGCCAAAT 1380
 Db 1321 ACAACACACCCCTATGCTGTTGCTTCAAAATCTCAGCATTAACATGAGTGGAGCCAAAT 1380
 QY 1381 TCAGAGCTTTCTCAAAATCAGATTAACATTCCTCAATTTCAATTAACAGGAGGAGGAGG 1440
 Db 1381 TCAGAGCTTTCTCAAAATCAGATTAACATTCCTCAATTTCAATTAACAGGAGGAGGAGG 1440

QY 1441 AGCCTAGAGTGCTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 Db 1441 AGCCTAGAGTGCTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 QY 1501 TGTCTCAACTCCCACTCCCACTCTGAGGAGTATTTCCCTCATTTCTACCTCTTAA 1560
 Db 1501 TGTCTCAACTCCCACTCCCACTCTGAGGAGTATTTCCCTCATTTCTACCTCTTAA 1560
 QY 1561 GGTATGACACCCCTCCCACTCCCACTCTGAGGAGTATTTCCCTCATTTCTACCTCTTAA 1620
 Db 1561 GGTATGACACCCCTCCCACTCCCACTCTGAGGAGTATTTCCCTCATTTCTACCTCTTAA 1620
 QY 1621 CCCATCTCCCATCTGGGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
 Db 1621 CCCATCTCCCATCTGGGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
 QY 1681 CCCCCAGTACTCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 1740
 Db 1681 CCCCCAGTACTCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 1740
 QY 1741 GTCTCAGTTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
 Db 1741 GTCTCAGTTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
 QY 1801 AGTGTGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
 Db 1801 AGTGTGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
 QY 1861 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
 Db 1861 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
 QY 1921 AGACNTCTGAGAGGCGCCTGCGCGAGGAGTCTCAAACTTAAAGGAGGAGGAGGAGGAGG 1980
 Db 1921 AGACNTCTGAGAGGCGCCTGCGCGAGGAGTCTCAAACTTAAAGGAGGAGGAGGAGGAGG 1980
 QY 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987
 RESULT 5
 ADI19771
 ID ADI19771 standard; DNA; 1987 BP.
 XX AC ADI19771;
 XX DT 22-APR-2004 (first entry)
 XX DE Human NOV4 DNA.
 XX KW Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 KW infection; anorexia; cancer; cardiovascular disease; hypertension;
 KW atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
 KW haemolipidic disorder; inflammatory skin disorder; asthma;
 KW dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 KW haemolipidic; wound healing; angiogenesis; chromosome mapping;
 KW tissue typing; preventive medicine; pharmacogenomic; gene therapy;
 KW anorectic; cardiant; vincidine; antibacterial; fungicide; protozoacide;
 KW neurotropic; neuroprotective; dermatological; human; gene; ds.
 OS Homo sapiens.
 XX PH Key
 XX FT Location/Qualifiers
 XX FT CDS .149
 XX FT /tag= b
 XX FT /product= "Human NOV protein"
 XX FT /transl_except= (pos:1171..1221, aa:Gly-Pro)
 XX FT /transl_except= (pos:1144..1152, aa:Val-Gly)
 XX FT sig_peptide
 XX FT /tag= a
 XX FT mat_peptide
 XX FT 1153..1446

FT /tag= C
 PN /product= "Human mature NOV protein"
 XX US2004002134-A1.
 XX 01-JAN-2004.
 XX
 PF 15-OCT-2001; 2001US-00977819.
 XX
 PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shinkets RA, Fernandes ER, Herrman JL, Vernet CM;
 XX WPI; 2004-070737/07.
 DR P-PSDB; ADI19772.
 XX
 PT New NOVX nucleic acids encoding human KIAA0768 protein-like and human
 PT protein PRO-228 polypeptides, useful for treating NOVX-associated
 PT disorders.
 XX
 PS Disclosure; SEQ ID NO 7; 95pp; English.
 XX
 CC The present invention is based in part on the discovery of novel secreted
 CC and membrane-bound polypeptides and their encoding polynucleotides. The
 CC nucleic acids and polypeptides are collectively referred as NOVX. The
 CC invention is useful for treating, preventing and diagnosing diseases such
 CC as metabolic disorders, diabetes, obesity, infectious diseases such as
 CC viral, bacterial, fungal, hemmthic and protozoal infections, anorexia,
 CC cancer, cardiovascular diseases such as hypertension and atherosclerosis,
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders such as osteoarthritis, haematopoietic
 CC disorders, inflammatory skin disorders, asthma and various dyslipidemias.
 CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV DNA.
 CC
 XX
 SQ Sequence 1987 BP; 448 A; 615 C; 530 G; 386 T; 0 U; 8 Other;
 Query Match 99.4%; Score 1975.8; DB 12; Length 1987;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1985; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 301 CCTACAGTATGTCGGGCAACGGCTCAACCCGCAATAGAAACAGTCCAGAGACCTTCCGGAC 360
 |||||
 Qy 361 GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCCTCCCGCCACGCTGC 420
 |||||
 Db 361 GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCCTCCCGCCACGCTGC 420
 |||||
 Qy 421 CAGCAAGGCTCCGGTAAACCCATTAACAAGGCTCCAGAGCATCTCAGCCAGGCTTCAGC 480
 |||||
 Db 421 CAGCAAGGCTCCGGTAAACCCATTAACAAGGCTCCAGAGCATCTCAGCCAGGCTTCAGC 480
 |||||
 Qy 481 CCCACGCAACCCCAAGACCCCAAGTGAACATCTCTCCCAATCAAGGCTCCCATCTATC 540
 |||||
 Db 481 CCCACGCAACCCCAAGACCCCAAGTGAACATCTCTCCCAATCAAGGCTCCCATCTATC 540
 |||||
 Qy 541 CAGGGCAGGCGCAGGAGTCTCCCTGGCTGACACATGATCCCAATTTCAATTTGGCTC 600
 |||||
 Db 541 CAGGGCAGGCGCAGGAGTCTCCCTGGCTGACACATGATCCCAATTTCAATTTGGCTC 600
 |||||
 Qy 601 CGTCACTTAATCCAGATACAGAGGCTGAGGAGGAGAGAGATCTAAAGAACCACT 660
 |||||
 Db 601 CGTCACTTAATCCAGATACAGAGGCTGAGGAGGAGAGAGATCTAAAGAACCACT 660
 |||||
 Qy 661 GTGGGTAGGGGAAATGGGACACAGCAGACATATGGGCTAGCTTCGACGACACAGACA 720
 |||||
 Db 661 GTGGGTAGGGGAAATGGGACACAGCAGACATATGGGCTAGCTTCGACGACACAGACA 720
 |||||
 Qy 721 GACAAACCTCTGATCTATGATAGTCTCTGACGGGCAAGGGGACAGGACCTTGAACCT 780
 |||||
 Db 721 GACAAACCTCTGATCTATGATAGTCTCTGACGGGCAAGGGGACAGGACCTTGAACCT 780
 |||||
 Qy 781 CTTGGCCAGGGGAGTGGGAGAGACAGAGGAAAGTTCACAGGCAAGGCTCTATCTAAG 840
 |||||
 Db 781 CTTGGCCAGGGGAGTGGGAGAGACAGAGGAAAGTTCACAGGCAAGGCTCTATCTAAG 840
 |||||
 Qy 841 TGGAACTAATTTGCCCGAGGGCTCAGCAAGGCCAAAGGAGACAGCCGTGACGTAACCT 900
 |||||
 Db 841 TGGAACTAATTTGCCCGAGGGCTCAGCAAGGCCAAAGGAGACAGCCGTGACGTAACCT 900
 |||||
 Qy 901 CCCCTTACAGGCTCTCCAGACCCCAAGCAGGAGGAGGCTGCTCCCAACCCGCTGCC 960
 |||||
 Db 901 CCCCTTACAGGCTCTCCAGACCCCAAGCAGGAGGAGGCTGCTCCCAACCCGCTGCC 960
 |||||
 Qy 961 CCAGCCAGCTGGCTGTGTCAGAGGAGGAGCCATGCAATGTATATAGTGGGTTT 1020
 |||||
 Db 961 CCAGCCAGCTGGCTGTGTCAGAGGAGGAGCCATGCAATGTATATAGTGGGTTT 1020
 |||||
 Qy 1021 CCAATACAGCTGGTGTGTAATAAATGATGAATCTCTGCGCTGCTGCTGGG 1080
 |||||
 Db 1021 CCAATACAGCTGGTGTGTAATAAATGATGAATCTCTGCGCTGCTGCTGGG 1080
 |||||
 Qy 1081 CTTCCAGGCAAGGCAAGTGGGCTGGGCTGGCTCTCTCTCTCCCAAGGCC 1140
 |||||
 Db 1081 CTTCCAGGCAAGGCAAGTGGGCTGGGCTGGCTCTCTCTCTCCCAAGGCC 1140
 |||||
 Qy 1141 TGTGTTCTGTTGGGCTGTCCCATGACAGACAGATCACTTACAGATGGAACCAAGGC 1200
 |||||
 Db 1141 TGTGTTCTGTTGGGCTGTCCCATGACAGACAGATCACTTACAGATGGAACCAAGGC 1200
 |||||
 Qy 1201 ATGATGAGGCTTTGGCTCTGAGGTTGAGACCCAGCTCTTCCACCTCCCTCCCG 1260
 |||||
 Db 1201 ATGATGAGGCTTTGGCTCTGAGGTTGAGACCCAGCTCTTCCACCTCCCTCCCG 1260
 |||||
 Qy 1261 CAGTCAGCTCTCCATCCATCCCTCTTTATCTATGAACTATATAGCTGAGTGTGTA 1320
 |||||
 Db 1261 CAGTCAGCTCTCCATCCATCCCTCTTTATCTATGAACTATATAGCTGAGTGTGTA 1320
 |||||
 Qy 1321 ACAACACACCCCTATGCTGCTCTCAAAATCTACATTTACATTTGTTAGGCCAAAT 1380
 |||||
 Db 1321 ACAACACACCCCTATGCTGCTCTCAAAATCTACATTTACATTTGTTAGGCCAAAT 1380
 |||||
 Qy 1381 TCAAGCTTCTCAATCAATTTACATTTCAATTTCAATTTCAATTTCAATTTCAATTT 1440
 |||||

Db 181 AGTGCCTGACGAGCACTGTACATGAAACCTTACGACTACAAAGTGCAGACT 240
 QY 241 GCCGAGAGAGTGAACCCCGAGCGTGCAGTCTGCTGAGAGAGAGAGAGAG 300
 Db 241 GCCGAGAGAGTGAACCCCGAGCGTGCAGTCTGCTGAGAGAGAGAGAGAG 300
 QY 301 CTTCACTAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 301 CTTCACTAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 361 GGAGATCAAAATGTAGTGCCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
 Db 361 GGAGATCAAAATGTAGTGCCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
 QY 421 CAGCAAGCTGCTTAACTCCATTTACAAAGCTCCAGAGAGAGAGAGAGAG 480
 Db 421 CAGCAAGCTGCTTAACTCCATTTACAAAGCTCCAGAGAGAGAGAGAGAG 480
 QY 481 CCCGAG 540
 Db 481 CCCGAG 540
 QY 541 CAGGAG 600
 Db 541 CAGGAG 600
 QY 601 CTTCACTAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 Db 601 CTTCACTAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 661 GTGGGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 Db 661 GTGGGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 721 GAGCAACCTCTCTATCTATGAGAGTCTCTGAGAGAGAGAGAGAGAGAG 780
 Db 721 GAGCAACCTCTCTATCTATGAGAGTCTCTGAGAGAGAGAGAGAGAGAG 780
 QY 781 CTTGAG 840
 Db 781 CTTGAG 840
 QY 841 TGGAGCTAATGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 Db 841 TGGAGCTAATGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 901 CCCCTGACAG 960
 Db 901 CCCCTGACAG 960
 QY 961 CAG 1020
 Db 961 CAG 1020
 QY 1021 CCAATCAAGTGTGTGTAAGAACTGCAATGAACTCTGCTGCTGCTGCTG 1080
 Db 1021 CCAATCAAGTGTGTGTAAGAACTGCAATGAACTCTGCTGCTGCTGCTG 1080
 QY 1081 CCTCAAG 1140
 Db 1081 CCTCAAG 1140
 QY 1141 TGTGTCNTGTGGGCTGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Db 1141 TGTGTCNTGTGGGCTGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1201 ATGAGATGAGGCTTGTGGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 Db 1201 ATGAGATGAGGCTTGTGGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1261 CAGTCAAGCTCTCCATCAATCCCTCTTTAATCTATGAACTATAGAGCTG 1320
 Db 1261 CAGTCAAGCTCTCCATCAATCCCTCTTTAATCTATGAACTATAGAGCTG 1320

QY 1321 ACAACACACCCCTATCGTGTCTCTCAATTAATCTAGCATTTACATTGTTGAGCCCAAT 1380
 Db 1321 ACAACACACCCCTATCGTGTCTCTCAATTAATCTAGCATTTACATTGTTGAGCCCAAT 1380
 QY 1381 TCAGAGCTTCTCAATTAATCAATTTACATCTCAATTTCTAATCAAGGAGAGAGAGAG 1440
 Db 1381 TCAGAGCTTCTCAATTAATCAATTTACATCTCAATTTCTAATCAAGGAGAGAGAGAG 1440
 QY 1441 AGCCACTGAGTGTGTGCTTTGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 Db 1441 AGCCACTGAGTGTGTGCTTTGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1501 TGTCTCAATCTCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Db 1501 TGTCTCAATCTCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 GGTATGACACCCCTCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Db 1561 GGTATGACACCCCTCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 CCCATCTCCATCTGAG 1680
 Db 1621 CCCATCTCCATCTGAG 1680
 QY 1681 CCCCAGTGAATCTGTTGCTCTCCATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 Db 1681 CCCCAGTGAATCTGTTGCTCTCCATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 QY 1741 GTCTCACTGAG 1800
 Db 1741 GTCTCACTGAG 1800
 QY 1801 AGTGTGCTTAAAGGTGAGCTCTCACTCCCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 Db 1801 AGTGTGCTTAAAGGTGAGCTCTCACTCCCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 1861 AGGAGGCTGAG 1920
 Db 1861 AGGAGGCTGAG 1920
 QY 1921 AGAATCTGAG 1980
 Db 1921 AGAATCTGAG 1980
 QY 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987
 RESULT 7
 AA231583
 ID AA231583 standard; cDNA; 2192 BP.
 XX
 AC AA231583;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Human breast tumour-associated EST 43.
 XX
 KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 XX
 OS Homo sapiens.
 XX
 PN DE19813835-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 20-MAR-1998; 98DE-01013835.
 XX
 PR 20-MAR-1998; 98DE-01013835.
 XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A,

XX WPI, 1999-528979/45.

DR P-PSDB; AAY48501, AAY48502.

XX Human nucleic acid sequences and protein products from normal breast
PT tissue, useful for breast cancer therapy.

XX Claim 1a; 131; 206bp; German.

XX This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytotstatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated with
CC fat metabolism. AA23541-235610 represent expressed sequence tags
CC described in the method of the invention

XX Sequence 2192 BP; 457 A; 717 C; 588 G; 430 T; 0 U; 0 Other;

Query Match 84.8%; Score 1684.4; DB 2; Length 2192;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1858; Conservative 0; Mismatches 73; Indels 24; Gaps 11;

Qy 33 GCCCATCCCAACCAAGCCCGTGGCGAGCAATCAAGCCTTAGAGACGCTTGA 92
Db 258 GGCCTTCCACGCCCCCGGTGGGAGCAATCAAGCCTTAGAGACGCTTGA 317
Qy 93 GTTTCGGTGAAGTGAAGATTCTCACTGAAGATCAATTGTCAACCTTCAACA 152
Db 318 GTTTCGGTGAAGTGAAGATTCTCACTGAAGATCAATTGTCAACCTTCAACA 374
Qy 153 AACCAATCTGAAGTGGCGGAGTGAAGAGTGGCGGTGAAGAGACATGATATAA 212
Db 375 CAACCAATCTGAAGTGGCGGAGTGAAGAGTGGCGGTGAAGAGACATGATATAA 432
Qy 213 CACCTTACGATCAACAAGTCCAGACTGCGGAGAGAGTGAAGCCTGAGCTGTGA 272
Db 433 TTGCT-----CACAAGTCCAG-CGCGGAGAGAGTGAAGCCTGAGCTGTGA 481
Qy 273 CCGTGGTACTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
Db 482 CCGTGG--CTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
Qy 333 CAGAAGAGTCCAGAGAGCTTCCGAGAGAGATCAAAATGTAGTGCCTTCCCTTCC 392
Db 539 CAGAAGAGTCCAGAGAGCTTCCGAGAGAGATCAAAATGTAGTGCCTTCCCTTCC 598
Qy 393 TTTCCTGT-CCCCCGCCAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 451
Db 599 TTTCCTGTGCCCCCGCCAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 658
Qy 452 CTCAGAGAGATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 511
Db 659 CTCAGAGAGATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 718
Qy 512 CCGTCCAACTAGAT 571
Db 719 CCGTCCAACTAGAT 778
Qy 572 CATGATGCCAGATTTTCAAGATTTTGGCTCCGTCACTTATCAAGATCAAGAGAG 631
Db 779 CATGATGCCAGATTTTCAAGATTTTGGCTCCGTCACTTATCAAGATCAAGAGAG 838
Qy 632 TCAGGAGAGAGAGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 691
Db 839 TCAGGAGAGAGAGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 898
Qy 692 ATGGGCAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 751

Db 899 ATGGGCAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 958
Qy 752 GGGCAAGGGAGACCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 811
Db 959 GGGCAAGGGAGACCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1018
Qy 812 AAGGTACAGAT 871
Db 1019 AAGGTACAGAT 1078
Qy 872 CAGAAGAT 931
Db 1079 CAGAAGAT 1138
Qy 932 CGAT 991
Db 1139 CGAT 1198
Qy 992 TGCCATCTGTATATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1051
Db 1199 TGCCATCTGTATATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1258
Qy 1052 AAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1111
Db 1259 AAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318
Qy 1112 GGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1171
Db 1319 GGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1378
Qy 1172 GATACCTTAACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1231
Db 1379 GATACCTTAACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1438
Qy 1232 CCGGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1290
Db 1439 CCGGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1498
Qy 1291 ATCTATGAATCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1349
Db 1499 ATCTATGAATCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1558
Qy 1350 TACTGAGATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1409
Db 1559 TACTGAGATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1618
Qy 1410 CTCATTTTCAATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1469
Db 1619 CTCATTTTCAATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1678
Qy 1470 AGGTATGATCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1529
Db 1679 AGGTATGATCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1738
Qy 1530 AGGATATTTTCCCTCATTTCTAAGAGAGAGAGAGAGAGAGAGAGAT 1589
Db 1739 AGGATATTTTCCCTCATTTCTAAGAGAGAGAGAGAGAGAGAGAGAT 1798
Qy 1590 CTGGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1649
Db 1799 CTGGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1858
Qy 1650 CATCAGCTTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1709
Db 1859 CATCAGCTTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1918
Qy 1710 CATACAGAT 1769
Db 1919 CATACAGAT 1978
Qy 1770 GTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1829

Db 1979 GTACGCTCTCCCTCCATCATCTCTCTGTCTGTCAGTGTGCTTAAAGTGGAGCTCTCACTCCC 2038
Qy 1830 ACCACTTTGGG-CCCTTGGGGGAGACTGGGGAGGGGGCCCTGGGAGAGCCCTGAGAGCTG 1888
Db 2039 ACCACTTGGAGCCCTTGGGGAGAGACTGGGGAGGGGGCCCTGGGAGAGCCCTGAGAGCTG 2098
Qy 1889 GAACTGTATATACATATMAAGACAGTCTCACAGACNTCTGAGAGCCGCTGGCCGAGT 1948
Db 2099 GAACCTGTATATACATATMAAGACAGTCTCACAGACMAAAGAGGCGCTGCTGCC-CGAGT 2157
Qy 1949 TCTCAAACTTAAAGGCGAGGCGNNNCTTACTTGAGA 1983
Db 2158 TCTCAAACTTAAAGGCGAGGCGCTTACTTGAGAGAAA 2192

RESULT 8
AA160565
ID AA160565 standard; cDNA; 2142 BP.
XX
AC AA160565;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4554.
XX
XX Human, neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocarcin;
XX chemotherapeutic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
XX
PR 21-JAN-2000; 2000US-00488725.
XX
PR 25-APR-2000; 2000US-00552317.
XX
PR 20-JUN-2000; 2000US-00598042.
XX
PR 19-JUL-2000; 2000US-00620312.
XX
PR 03-AUG-2000; 2000US-00653450.
XX
PR 14-SEP-2000; 2000US-00662191.
XX
PR 19-OCT-2000; 2000US-00693036.
XX
PR 29-NOV-2000; 2000US-00727344.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
XX Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,
XX Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX
XX P-PSDB; AAM41409.
XX
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX
XX Claim 1; SEQ ID NO 4554; 10078bp; English.
XX
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
XX encoded polypeptides (AAM38642-AA42213) with neotropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotherapeutic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other;
Query Match 84.0%; Score 1670; DB 4; Length 2142;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1815; Conservative 0; Mismatches 57; Indels 22; Gaps 9;
Qy 33 GCCATCCCAACCAAGCCCGGTGGGAGGAGCAATCAAGACCTTAGAGAGCGCTTGA 92
Db 259 GGCCTCCCAAGCCCGGCGCGGTGGGAGGAGCAATCAAGACCTTAGAGAGCGCTTGA 318
Qy 93 GTTTCGGGTGAGCTTGAAGAGCTTCTCACTTAAAGACATATTGTACACCTTCAACA 152
Db 319 GTTTCGGGTGAGCTTGAAGAGCTTCTCACTTAAAGACATATTGTACACCTTCAACA 375
Qy 153 AACCAATCAGAGTGGGGGGCTGAGAAAGCTGGCGCTGAGAGGACATGTACATGAAA 212
Db 376 CAACCAATCAGAGTGGGGGGCTGAGAAAGCTGGCGCTGAGAGGACATGTACATGAAA 433
Qy 213 CACCTTACAGACTACAAAGTGCACAGCTGCGGAGACGTGAGACCCCGAGCGGTGA 272
Db 434 TTCCCT-----CACAGTGCAG--CTGCCGAGAGACGTGAGACCC--GACGTGGTGA 482
Qy 273 CCTGCGCTACTGTGCGGAGAGAGAGAGAGCTTCACTGTCCGGGACAGCGTCAACCCGAT 332
Db 483 CTTGGG--CTTGGGAGAGAGAGAGAGAGCTTCACTGTCCGGGACAGCGTCAACCCGAT 539
Qy 333 CAGAACAGTCCAGACAGACCTTCCGAGCGAGATCAAAATCTAGTGCCTTCCCTCC 392
Db 540 CAGAACAGTCCAGACAGACCTTCCGAGCGAGATCAAAATCTAGTGCCTTCCCTCC 599
Qy 393 TTTCCCTGTCCCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452
Db 600 TTTCCCTGTCCCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659
Qy 453 TCCAGACATCTCAGCCAGCGGCTTCAAGCCCGACAGACCCAGACCCAGTGAACATC 512
Db 660 TCCAGACATCTCAGCCAGCGGCTTCAAGCCCGACAGACCCAGACCCAGTGAACATC 719
Qy 513 CTCCCAACTAGAGGCGCTTCACTCTTATCCAGGCGAGGCGGAGCTTCCGCTTGAAC 572
Db 720 CTCCCAACTAGAGGCGCTTCACTCTTATCCAGGCGAGGCGGAGCTTCCGCTTGAAC 779
Qy 573 ATGATGCCAGATTTCAGATTGAGCTTGGCGCTGCACTTATCAGAGTACAGGGGCTGGGT 632
Db 780 ATGATGCCAGATTTCAGATTGAGCTTGGCGCTGCACTTATCAGAGTACAGGGGCTGGGT 839
Qy 633 CAGGAAAGAAAGTCTTAAAGACCCAGTGTGCTCAGGGGAAATGGAGCCAGAGACATA 692
Db 840 CAGGAAAGAAAGTCTTAAAGACCCAGTGTGCTCAGGGGAAATGGAGCCAGAGACATA 899
Qy 693 TGGGCAAGCTCTCAG 752
Db 900 TGGGCAAGCTCTCAG 959
Qy 753 GGCAGAGGAGACAGAGAGCTGGAACCTCTTGGCCAAAGGAGAGAGAGAGAGAGAGAG 812
Db 960 GGCAGAGGAGACAGAGAGCTGGAACCTCTTGGCCAAAGGAGAGAGAGAGAGAGAGAG 1019
Qy 813 AGGTACAGGACAG 872
Db 1020 AGGTACAGGACAG 1079
Qy 873 AAG 932
Db 1080 AAG 1139

XX Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other;
 SQ Query Match 84.0%; Score 1670; DB 10; Length 2142;
 Best Local Similarity 95.8%; Pred. No. 0;
 Matches 1815; Conservative 0; Mismatches 57; Indels 22; Gaps 9;

33 GCCCATCCCAACCAACCCCGTGGGCAAGCAACATCAAGACCTTGAAGAGCCCTATGA 92
 Db 259 GGGCTTCCCAACCCCGGCGGTGGGCAAGCAACATCAAGACCTTGAAGAGCCCTATGA 318
 Qy 93 GTTTCGGGTGAGAGTGAAGACTTCTCACTTGAAGATCATTTGTGACCACTCCAAACA 152
 Db 319 GTTTCGGGTGAGAGTGAAGACTTCTCACTTGAAGATCATTTGTGACCACTCCCA---A 375
 Qy 153 AACCAATCGAGGTGGGAGTGAAGAAAGTGGGCTGAAGAGCAAGCACTGTACATGAAA 212
 Db 376 CAACCAATCGAGGTGGGAG--CTGAAGAGTGGGCTGAAGAGCAAGCACTGTACATGAAA 433
 Qy 213 CACCTTACGACTACAAAGTGCAGACTGCGGAGAGAGTGAAGAGCCCGAGCGTGGGTGA 272
 Db 434 TTGGCT-----CACAAAGTGCAG--CTGCGGAGAGAGTGAAGCC---GAGGTGGTGA 482
 Qy 273 CTTCCGCTACTGTGGGAGAGAGAGAGAGCTCACTACTGCGGAGAGCGCTCAAGCGGATA 332
 Db 483 CTTGG---CTTGGGAGAGAGAGAGAGCTCACTACTGCGGAGAGCGCGTCAAGCGGATA 539
 Qy 333 CAGAAACGTCAGAGAGAGCTTCCGAGAGAGATCAAAAATGAGAGCTTCCCTCCCTCC 392
 Db 540 CAGAAACGTCAGAGAGAGCTTCCGAGAGAGATCAAAAATGAGAGCTTCCCTCCCTCC 599
 Qy 393 TTTTCCTGTCCCGCCCGCCAGCGCTGCGAGCAAGAGCTTGGTAACTCCCATTAACAAGC 452
 Db 600 TTTTCCTGTCCCGCCCGCCAGCGCTGCGAGCAAGAGCTTGGTAACTCCCATTAACAAGC 659
 Qy 453 TCCAGAGCATCTCAGAGAGAGTCTAGCCCGCCAGAGAGAGAGAGAGAGAGAGAGATC 512
 Db 660 TCCAGAGCATCTCAGAGAGAGTCTAGCCCGCCAGAGAGAGAGAGAGAGAGAGAGATC 719
 Qy 513 CTCCTCAAATGAGGAGCTTCCATCTTATCGAGGAGAGAGAGAGAGAGAGAGAGAGAG 572
 Db 720 CTCCTCAAATGAGGAGCTTCCATCTTATCGAGGAGAGAGAGAGAGAGAGAGAGAGAG 779
 Qy 573 ATGATGCCAGATTTGAGATTTGGCTCCGCTCACTTATCCAGAGAGAGAGAGAGAGAGAG 632
 Db 780 ATGATGCCAGATTTGAGATTTGGCTCCGCTCACTTATCCAGAGAGAGAGAGAGAGAGAG 839
 Qy 633 CAGGAGAGAGAGATCTTAAAGAACCACTGTGGGTGAGGAGAGAGAGAGAGAGAGAGAT 692
 Db 840 CAGGAGAGAGAGATCTTAAAGAACCACTGTGGGTGAGGAGAGAGAGAGAGAGAGAGAT 899
 Qy 693 TGGGCAAGCTCTGCAAG 752
 Db 900 TGGGCAAGCTCTGCAAG 959
 Qy 753 GGCAG 812
 Db 960 GGCAG 1019
 Qy 813 AGGTTCAG 872
 Db 1020 AGGTTCAG 1079
 Qy 873 AAG 932
 Db 1080 AAG 1139
 Qy 933 GAGCAG 992
 Db 1140 GAGCAG 1199
 Qy 993 GGCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTCTGTGAAAATCTGATGA 1052

Db 1200 GGCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTCTGTGAAAATCTGATGA 1259
 Qy 1053 AACTCTGCGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTG 1112
 Db 1260 AACTCTGCGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTCTGCGGCTG 1319
 Qy 1113 GGGCTGCT 1172
 Db 1320 GGGCTGCT 1379
 Qy 1173 ATCAGCTTAAAGAGATGAG 1232
 Db 1380 ATCAGCTTAAAGAGATGAG 1439
 Qy 1233 CCGAGCTTCTGCGGCT 1291
 Db 1440 CCGAGCTTCTGCGGCT 1499
 Qy 1292 TCTATGATCTATAGAGTGGTGTGTGTAA--ACACACCTTATCGTTGCTCTTCAAT 1350
 Db 1500 TCTATGATCTATAGAGTGGTGTGTGTAA--ACACACCTTATCGTTGCTCTTCAAT 1559
 Qy 1351 ACTCAGCATTTACATTTGGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410
 Db 1560 ACTCAGCATTTACATTTGGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
 Qy 1411 TCCATTTTCAATTAAGGAG 1470
 Db 1620 TCCATTTTCAATTAAGGAG 1579
 Qy 1471 GGTTAGATCTGAG 1530
 Db 1680 GGTTAGATCTGAG 1739
 Qy 1531 GGAATATTTCCCTCATTTCTACCTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1590
 Db 1740 GGAATATTTCCCTCATTTCTACCTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1799
 Qy 1591 TGGGAGATGGGAG 1650
 Db 1800 TGGGAGATGGGAG 1859
 Qy 1651 ATCAGCTTAAAGCTTGGGAG 1710
 Db 1860 ATCAGCTTAAAGCTTGGGAG 1919
 Qy 1711 ATCAG 1770
 Db 1920 ATCAG 1979
 Qy 1771 TCAAGCT 1830
 Db 1980 TCAAGCT 2039
 Qy 1831 CCACTCTGAG--CCCTTGGGAG 1889
 Db 2040 CCACTCTGAG--CCCTTGGGAG 2099
 Qy 1890 AACCTGTATACAAATTAAG 1923
 Db 2100 AACCTGTATACAAATTAAG 2133

RESULT 10
 ADM1.9231
 ID ADM1.9231 standard; cdNA; 2220 BP.
 XX ADM1.9231;
 AC ADM1.9231;
 XX 20-MAY-2004 (first entry)
 DT
 XX
 DE Novel human channel/transporter gene #28.

KW de; gene; immunosuppressive; antiarthritic; anti-rheumatic;
KW antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; gene therapy; channel/transporter protein;
KW rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW cerebral ischemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
XX
OS Homo sapiens.
XX
XX WO200154472-A2.
XX
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205555P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217486P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227909P.
PR 30-AUG-2000; 2000US-0228928P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234938P.
PR 26-SEP-2000; 2000US-0234944P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0244675P.
PR 08-NOV-2000; 2000US-0244676P.
PR 08-NOV-2000; 2000US-0244677P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246612P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249248P.
PR 17-NOV-2000; 2000US-0249249P.
PR 17-NOV-2000; 2000US-0249250P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.

PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
DR P-PSDB; ADM19710.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
XX used in preventing, treating or ameliorating a medical condition.
PS Claim 1; SEQ ID NO 38; 809pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a gene of the
CC invention.
XX
SQ Sequence 2220 BP; 461 A; 718 C; 597 G; 429 T; 0 U; 15 Other;
Query Match 83.9%; Score 1668; DB 5; Length 2220;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1810; Conservative 5; Mismatches 57; Indels 22; Gaps 9;
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGCAATCAAGACCTTAGAGAGCGCTATGA 92
DB 313 GGGCTTCCAGCCCGCCCGGTGGGAGCAATCAAGACCTTAGAGAGCGCTATGA 372
QY 93 GTTTGGGTGGAGCGTGAAGACTTTCACTGAAGACATCATTTGACCACTTCAACA 152
DB 373 GTTTGGGTGGAGCGTGAAGACTTTCACTGAAGACATCATTTGACCACTTCAACA 429
QY 153 AACCAATCGAGTGGCGGGCTGAGAAAGCTGGCGGCTGAACGACACTGATCAAGAA 212
DB 430 CAACCAATCGAGTGGCGGGCTGAGAAAGCTGGCGGCTGAACGACACTGATCAAGAA 487
QY 213 CACCTTACGACTCAACAAAGTGCAGACTGCGGAGAGCGTGAACCCGAGCGTGGTA 272
DB 488 TTGGCT-----CAAGAATGCGAG-CTGCCGAGAGAGCTGAGACC--GAGCTGGTGA 536
QY 273 CTTGGGCTACTGTGGGAGAGAGCGAGCAAGCTCACTAGTCCGGGAGCGGTGATACCCGATA 332
DB 537 CTTGGG---CTTGGGAGAGAGCGAGCAAGCTCACTAGTCCGGGAGCGGTGATACCCGATA 593
QY 333 CAGAAACAGCTCCAGAGACTTTCGAGAGAGATCAAAATCTGATGCTCTCCCTTCCC 392
DB 594 CAGAACAGCTCCAGAGACTTTCGAGAGAGATCAAAATCTGATGCTCTCCCTTCCC 653

QY 393 TTTCCTGTGCCCCCGCCCGCCAGCGCTGCAGCAAAAGCTGCTTAACCCATTACAAGC 452
DB 654 TTTCCTGTGCCCCCGCCCGCCAGCGCTGCAGCAAAAGCTGCTTAACCCATTACAAGC 713
QY 453 TCGAGCATCTCAGCCCGAGTTCTAGCCCGCCAGCGACCCCAAGCCCCAGGTGACCATC 512
DB 714 TCGAGCATCTCAGCCCGAGTTCTAGCCCGCCAGCGACCCCAAGCCCCAGGTGACCATC 773
QY 513 CTCCCAAACTAGGGGCGCTTCACTATTCAGGGGAGAGGCACTCCCTGGCTGACAC 572
DB 774 CTCCCAAACTAGGGGCGCTTCACTATTCAGGGGAGAGGCACTCCCTGGCTGACAC 833
QY 573 ATGATGCCAGATTTCAGATTGAGCTCCCTGCTCACTTATCAGAGTACAGGGGCTGGGCT 632
DB 834 ATGATGCCAGATTTCAGATTGAGCTCCCTGCTCACTTATCAGAGTACAGGGGCTGGGCT 893
QY 633 CAGGAAAGAAAGATCTTAAGAACCACTAGTGGGCTAGGGGAATGGGACCAAGACATTA 692
DB 894 CAGGAAAGAAAGATCTTAAGAACCACTAGTGGGCTAGGGGAATGGGACCAAGACATTA 953
QY 693 TGGCAAGCTCTGAGAGACAGACAGACAGACAAACCTGATCTATGAAGTCTGCGAG 752
DB 954 TGGCAAGCTCTGAGAGACAGACAGACAGACAAACCTGATCTATGAAGTCTGCGAG 1013
QY 753 GGCAGGGGACACAGGACCTGGAACCTCTTGGCCAAAGGGAGTGGAGAGACAGAGGA 812
DB 1014 GGCAGGGGACACAGGACCTGGAACCTCTTGGCCAAAGGGAGTGGAGAGACAGAGGA 1073
QY 813 AGGTCAAGGCAAGGGGCTTATCTAGTGAATTAATGGCCGAGGCTCAGCAAGGCG 872
DB 1074 AGGTCAAGGCAAGGGGCTTATCTAGTGAATTAATGGCCGAGGCTCAGCAAGGCG 1133
QY 873 AAGAGAGACAGCGCTGACGCTAACTCCCTCTACAGACCTCCAGGCCACAGCGCAGC 932
DB 1134 AAGAGAGACAGCGCTGACGCTAACTCCCTCTACAGACCTCCAGGCCACAGCGCAGC 1193
QY 933 GAGAGAGCTGCGCCCAACCCCGTGGCCCAAGCACTGCTGTGCGAGGAGACCAT 992
DB 1194 GAGAGAGCTGCGCCCAACCCCGTGGCCCAAGCACTGCTGTGCGAGGAGACCAT 1253
QY 993 GCCACATCTGTATATAGATGGGTTTTCCTCAATACAGCTGTTGCGAATAAATCGATGA 1052
DB 1254 GCCACATCTGTATATAGATGGGTTTTCCTCAATACAGCTGTTGCGAATAAATCGATGA 1313
QY 1053 AACTCTGCGCTCTGCGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1112
DB 1314 AACTCTGCGCTCTGCGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1373
QY 1113 GGGCTGCTCTTCTCCCTCCACAGGCTGTGTGTTGCGGCTGCTCCATGACAGCAGG 1172
DB 1374 GGGCTGCTCTTCTCCCTCCACAGGCTGTGTGTTGCGGCTGCTCCATGACAGCAGG 1433
QY 1173 ATCACTTAACAGATGAGAGCCAGGGCATGATGGGCTTTGGTCTCGAGGTTGAGC 1232
DB 1434 ATCACTTAACAGATGAGAGCCAGGGCATGATGGGCTTTGGTCTCGAGGTTGAGC 1493
QY 1233 CCCAGCTCTGCGCACTTCCCTCCCTC--GGCACTGAGCTGCTCACTATCCCTCTTAA 1291
DB 1494 CCCAGCTCTGCGCACTTCCCTCCCTC--GGCACTGAGCTGCTCACTATCCCTCTTAA 1553
QY 1292 TCTATGAATCTATAGCTGAGTGTGTGAACA--AGACACCCCTATGTTGCTCTCAAT 1350
DB 1554 TCTATGAATCTATAGCTGAGTGTGTGAACA--AGACACCCCTATGTTGCTCTCAAT 1613
QY 1351 ACTCAGCATTTACATTTGTTGAGGCCAAATTAGAGCTTTCTCAATCAGATTTCAATC 1410
DB 1614 ACTCAGCATTTACATTTGTTGAGGCCAAATTAGAGCTTTCTCAATCAGATTTCAATC 1673
QY 1411 TCCATTTTCAATTAACGGGAAACATTCGCCAGACCTAGTGTGCTGCTTGTGCTGCTGAA 1470
DB 1674 TCCATTTTCAATTAACGGGAAACATTCGCCAGACCTAGTGTGCTGCTTGTGCTGCTGAA 1733

QY 1471 GGTTAGTCTGAACCCAGGGGTGTCACACGCTGCTCACTCCCACTCTGGGCACTGA 1530
Db 1734 GGTTAGTCTGAACCCAGGGGTGTCACACGCTGCTCACTCCCACTCTGGGCACTGA 1793
QY 1531 GGAGTATTTCCCTCAATTTCTAAGGCTATGCAACCCCTCCCAAGTCTTTCAGC 1590
Db 1794 GGAGTATTTCCCTCAATTTCTAAGGCTATGCAACCCCTCCCAAGTCTTTCAGC 1853
QY 1591 TGGGGATGGGGAGTCATAGGAAAAGCCCATCTCCCATCTGGGATAGGGAACCTTCC 1650
Db 1854 TGGGGATGGGGAGTCATAGGAAAAGCCCATCTCCCATCTGGGATAGGGAACCTTCC 1913
QY 1651 ATCAGCTTAACCTGGGAATGCTGCTGCCCCCAGTGAAGTCTTGTTCCTCCAC 1710
Db 1914 ATCAGCTTAACCTGGGAATGCTGCTGCCCCCAGTGAAGTCTTGTTCCTCCAC 1973
QY 1711 ATACAGAGCAGGCTGAGGGGAGAGGCTGCTCACTTACAGGGGTCTCCAGGGCAAG 1770
Db 1974 ATACAGAGCAGGCTGAGGGGAGAGGCTGCTCACTTACAGGGGTCTCCAGGGCAAG 2033
QY 1771 TCAGCCTCCTCCCTCCATGCTCTGTGAGTGTGCTTAAAGGCTCTCACTCCCA 1830
Db 2034 TCAGCCTCCTCCCTCCATGCTCTGTGAGTGTGCTTAAAGGCTCTCACTCCCA 2093
QY 1831 CCACTCTGGG-CCCTTGGGGAGAGCTGGGAGGGGGCGTGGAGAGCCCTGACCTGG 1889
Db 2094 CCACTCTGGGCGCCCTTGGGGAGAGCTGGGAGGGGGCGTGGAGAGCCCTGACCTGG 2153
QY 1890 AACCTGTATACAAATTAAGACAGTCTCACAGA 1923
Db 2154 AACCTGTATACAAATTAAGACAGTCTCACAGA 2187

RESULT 11
ADQ22267
ID ADQ22267 standard; DNA; 3503 BP.
AC ADQ22267;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5087.
XX
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.
XX Homo sapiens.
XX OS
XX PN WC0204048938-A2.
XX PD 10-JUN-2004.
XX PF 26-NOV-2003; 2003WO-US038193.
XX PR 26-NOV-2002; 2002US-0429739P.
XX (PROT-) PROTEIN DESIGN LABS INC.
XX PI Aziz N, Ginsburg WM, Zlotnik A;
XX MPI; 2004-441208/41.
XX DR
XX PT Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX PT and comparing the gene expression, also useful in treating soft tissue
XX PT sarcoma.
XX PS
XX Example 2; SEQ ID NO 5087; 210BP; English.
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 3503 BP; 815 A; 1068 C; 977 G; 642 T; 0 U; 1 Other;
Query Match 83.9%; Score 1666.8; DB 12; Length 3503;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1813; Conservative 0; Mismatches 59; Indels 22; Gaps 9;
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGAGCAATCAAGACCTTAGAGAGCCTATGA 92
Db 1371 GGCCTTCCAGCCCGCCCGGTGGGAGAGCAATCAAGACCTTAGAGAGCCTATGA 1430
QY 93 GTTTGCGGTGACGTGAGAGACTTCTCACTGAAGACATCATTTGTACACCTCGAACA 152
Db 1431 GTTTGCGGTGACGTGAGAGACTTCTCACTGAAGACATCATTTGTACACCTCGA--A 1487
QY 153 AACCAACATCGAGGTGGGGGCTGAGAGAGCTGGGGGCTGAACGGAGACTGTATGATAA 212
Db 1488 CAACCAACATCGAGGTGGGGG--CTGAGAGAGCTGGGGGCTGAACGGAGACTGTATGATAA 1545
QY 213 CACCTTAGACATCAAGTGCACAGTCCGAGAGAGAGTGGAGCCCGAGCGTGGGTGA 272
Db 1546 TTGGCT-----CACAAGTGCAG-CTGCCGAGAGAGCTGGAGCC--GACGTGGGTGA 1594
QY 273 CCTGGCTACTGTGGGAGAGAGAGAGAGCTTCACTAGTCCGGGAGAGGTCAACCCGATA 332
Db 1595 CCTGGG---CTGTGGGAGAGAGAGAGAGCTTCACTAGTCCGGGAGAGGTCAACCCGATA 1651
QY 333 CAGAACAGTCCAGAGAGACTTCCGAGAGAGATCAAAATCTGAGTGCCTCTCCCTTCC 392
Db 1652 CAGAACAGTCCAGAGAGACTTCCGAGAGAGATCAAAATCTGAGTGCCTCTCCCTTCC 1711
QY 393 TTTCCTGTCCCGCCGCCAGCGCTGTCAGCAAAAGCTGTGTAACCCCATTTACAAGC 452
Db 1712 TTTCCTGTCCCGCCGCCAGCGCTGTCAGCAAAAGCTGTGTAACCCCATTTACAAGC 1771
QY 453 TCCAGAGACATTCACACCCAGAGTTCTAGGCCAGAGCAACCCAGAGGTGAGCATC 512
Db 1772 TCCAGAGACATTCACACCCAGAGTTCTAGGCCAGAGCAACCCAGAGGTGAGCATC 1831
QY 513 CTCCAAACTAGAGGCGCTTCACTCTATCCAGGGAGCCAGGAGTCCCTGGCTGAGC 572
Db 1832 CTCCAAACTAGAGGCGCTTCACTCTATCCAGGGAGCCAGGAGTCCCTGGCTGAGC 1891
QY 573 ATGATGCCAGATTTTCAAGTTTGGCTCGCTACTTTATCCAGATTAAGAGGGCTGGGT 632
Db 1892 ATGATGCCAGATTTTCAAGTTTGGCTCGCTACTTTATCCAGATTAAGAGGGCTGGGT 1951
QY 633 CAGGAGAGAGATCTAAAGAACCACTGTGGGTGAGGGATGGGACAGAGACATA 692
Db 1952 CAGGAGAGAGATCTAAAGAACCACTGTGGGTGAGGGATGGGACAGAGACATA 2011
QY 693 TGGGCAAGCTCTGACAGAGACAGACAGACAAACCTCTTATATGAGTCTTGCAG 752
Db 2012 TGGGCAAGCTCTGACAGAGACAGACAGACAAACCTCTTATATGAGTCTTGCAG 2071
QY 753 GGTAAAGGAGCAGAGGAGCTGGAACCTTTGGCCAAAGGGAGTGGAGAGACAGAGGA 812
Db 2072 GGTAAAGGAGCAGAGGAGCTGGAACCTTTGGCCAAAGGGAGTGGAGAGACAGAGGA 2131
QY 813 AGGTCAAGGAGAGGTGCTTATCTTAAGTGAATTTGCCGAGAGGCTCAGCAAGGC 872
Db 2132 AGGTCAAGGAGAGGTGCTTATCTTAAGTGAATTTGCCGAGAGGCTCAGCAAGGC 2191
QY 873 AAGAGAGAGAGCGGTGAGTAACTTCCCTTACAGAGCTCCAAAGCCCAAGCCGAGC 932

[illegible]

RESULT 12

ID	ADB62818
AD	ADB62818 standard; CDNA; 2410 BP.
AC	
XX	ADB62818;
DT	
XX	
DE	04-DEC-2003 (first entry)
XX	
DB	Human cDNA encoding clone NTONG20031580.
XX	
KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KM	tissue regeneration; cell regeneration; membrane protein;
KW	signal transduction-related protein; transcription-related
KW	osteoporosis; neurological disease; cancer; tumour.
OS	
XX	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	126..863
CDS	/tag=a
FT	/product= "Clone NTONG20031580 protein"
FT	
PN	EPI308459-A2.
XX	
PD	07-MAY-2003.
XX	
PF	28-MAR-2002; 2002EP-00007401.
PR	
PR	05-NOV-2001; 2001JP-00379298.
PR	25-JAN-2002; 2002US-00350978.
XX	
PA	(HELI-) HELIX RES INST.
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.
PI	Isoegal T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Iehii S;
PI	Yanamoto J, Isoho Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeuno Y;
DR	WPJ, 2003-450961/43.
DR	P-PDSB; ADB64788.
XX	
PT	New polynucleotides and polypeptides, useful for developing a diagnostic
PT	marker or medicines for regulation of their expression and activity, or
PT	as targets of gene therapy.
XX	
PS	Claim 1; Page; 222pp; English.
XX	
CC	The invention discloses a polynucleotide comprising a sequence selected
CC	from 1970 fully defined nucleotide sequences which encode novel
CC	polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC	or its partial peptide, an antibody binding to the polypeptide or peptide
CC	of the polynucleotide, immunologically assaying the polypeptide or
CC	peptide of the polynucleotide by contacting the polypeptide or peptide
CC	with the antibody of the encoded protein, and observing the binding
CC	between the two, a transformant carrying the polynucleotide in an
CC	expressible manner and an antisense polynucleotide. The oligonucleotide
CC	is useful as a primer for synthesizing the polynucleotide, or as a probe
CC	for detecting the polynucleotide. The polynucleotides and encoded
CC	proteins are useful as pharmaceutical agents and many disease-related
CC	genes may be included in them, for developing a diagnostic marker or
CC	medicines for regulation of their expression and activity, or as targets
CC	of gene therapy. The genes are involved in tissue and/or cell
CC	regeneration. Membrane proteins, signal transduction-related protein,
CC	transcription-related proteins, disease-related proteins and genes
CC	encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC	neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC	the activity or expression of the encoded protein to treat diseases. The
CC	sequence presented is a cDNA of the invention. Note: Some of the sequence
CC	data for this patent is not represented in the printed specification, but
CC	is based on sequence information supplied by the European Patent Office.
XX	
SQ	Sequence 2410 BP; 538 A; 737 C; 680 G; 455 T; 0 U; 0 Other;

Query Match 83.6%; Score 1660.6; DB 10; Length 2410;

Best Local Similarity 95.6%; Pred. No. 0;
Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

```
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGGCAATCAAGACCTTAGAGACGCTATGA 92
Db 539 GGCCTTCCAGCCCGCCCGGTGGGAGGCAATCAAGACCTTAGAGACGCTATGA 598
QY 93 GTTGGGTGAGAGTGAAGATCTTCACTGAAGACATCATGTATCAACACCTCCAAAGA 152
Db 599 GTTGGGTGAGAGTGAAGATCTTCACTGAAGACATCATGTATCAACACCTCCAAAGA 655
QY 153 AACCAACATGAGGTGGGAGTGAAGAAAGCTGGCGCTGAACGGAACATGTATCATGAA 212
Db 656 CAACCAATGAGGTGGGAGTGAAGAAAGCTGGCGCTGAACGGAACATGTATCATGAA 713
QY 213 CACTTACGACTACAAAGTCCAGACTGCGGAGAGCTGGAACCCCGAGCGTGGGTGA 272
Db 714 TTGGCT-----CACAAAGTCCAG-CTGCGGAGAGCTGGAACCC--GACGTGGTGA 762
QY 273 CCTGGCTACTGTGGGAGAGGAGACGACTCACTAGTCCGGGACGCGTCAACCCGCAATA 332
Db 763 CTTGG---CTGTGGGAGAGAGCGGACCTCACTATCCGGGACGCGCTCAACCCGCAATA 819
QY 333 CAGAACGCTCCAGACAGACTCTTCGAGCGAGATCAAAATCTGATGCTCTCCCTTCCC 392
Db 820 CAGAACGCTCCAGACAGACTCTTCGAGCGAGATCAAAATCTGATGCTCTCCCTTCCC 879
QY 393 TTTTCCCTGTCCCGCCCGCCAGCGCTCCGAGCAAAAGCTGCTAAACCCCTTCAACAGC 452
Db 880 TTTTCCCTGTCCCGCCCGCCAGCGCTCCGAGCAAAAGCTGCTTAAACCCCTTCAACAGC 939
QY 453 TCCAGGACATCTCAGCCAGGTTCTAGCCCCCAACGCAACCCAGACCCAGGTGAGACATC 512
Db 940 TCCAGGACATCTCAGCCAGGTTCTAGCCCCCAACGCAACCCAGGTGAGACATC 999
QY 513 CTCCCAACTAGGAGCCCTCCACTTATCCAGGGAGAGGCCAGGACTCCCTGGGCTGAC 572
Db 1000 CTCCCAACTAGGAGCCCTCCACTTATCCAGGGAGAGGCCAGGACTCCCTGGGCTGAC 1059
QY 573 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTCAAGGGCTGGGGT 632
Db 1060 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTCAAGGGCTGGGGT 1119
QY 633 CAGGGAGAAAGATCTTAAAGAACCCACTGTGGGTCAAGGGAAATGGGACCAAGGACATA 692
Db 1120 CAGGGAGAAAGATCTTAAAGAACCCACTGTGGGTCAAGGGAAATGGGACCAAGGACATA 1179
QY 693 TGGGCAAGCTCTGAGGACAGACAGACAGACAAACCTGTGATCTATAGAGTCTCTGAC 752
Db 1180 TGGGCAAGCTCTGAGGACAGACAGACAGACAAACCTGTGATCTATAGAGTCTCTGAC 1239
QY 753 GGCAGAGGAGACCAAGGACCTTGAACCTCTTGGCCAAAGGAGAGTGGAGAGACAGAGGA 812
Db 1240 GGCAGAGGAGACCAAGGACCTTGAACCTCTTGGCCAAAGGAGAGTGGAGAGACAGAGGA 1299
QY 813 AGGTCAAGGAGAAAGGTGCTTATAGTGGAACTAATTGCCGAGGGGCTCAGAGAGGC 872
Db 1300 AGGTCAAGGAGAAAGGTGCTTATAGTGGAACTAATTGCCGAGGGGCTCAGAGAGGC 1359
QY 873 AAGAGGAGACAGCCGTGACGTTAACTTCCCTCTACAGAGCTTCAAGCCCAAGCCAGC 932
Db 1360 AAGAGGAGACAGCCGTGAGTTAACTTCCCTCTACAGAGCTTCAAGCCCAAGCCAGC 1419
QY 933 GAGCAGGTGCTGCTCCCAACCCCGTGGCCCAAGCAGCTGCTGTCACAGGACAGCCAT 992
Db 1420 GAGCAGGTGCTGCTCCCAACCCCGTGGCCCAAGCAGCTGCTGTCACAGGACAGCCAT 1479
QY 993 GCCACATCTGATATAGATGGGGTTTTTCCAAATCAGAGTGGTGTGMAAAACTGCATGA 1052
Db 1480 GCCACATCTGATATAGATGGGGTTTTTCCAAATCAGAGTGGTGTGATAAACTGCATGA 1539
QY 1053 AACTCTGCGGTCTGCGGCTGCTGGGGCTTCCAGGCAAGGCCAAAGTGGGGTTGGGGGTG 1112
```

```
Db 1540 AACTCTGCGGTCTGCGGCTGCTGGGGCTTCCAGGCAAGGCCAAGTGGGGTTGGGTGTG 1599
QY 1113 GGGCTGTGCTCTTCTCCCTCCCAAGGCTGTGTTCTAGGGGCTGTCCCAAGCAGACAG 1112
Db 1600 GGGCTGTGCTCTTCTCCCTCCCAAGGCTGTGTTCTAGGGGCTGTCCCAAGCAGACAG 1659
QY 1173 ATCACTTAACAGAGATGAAAGCCAGGAGATGAGGCTTTGGGCTCTCGAGGTGAC 1232
Db 1660 ATCACTTAACAGAGATGAAAGCCAGGAGATGAGGCTTTGGGCTCTCGAGGTGAC 1719
QY 1233 CCGAGCTTGTGCCACTTCCCTGCC--GGAGTCAAGCTCTGCATCCATCCCTCTTTAA 1291
Db 1720 CCGAGCTTGTGCCACTTCCCTGCCGAGTCAAGCTCTGCATCCATCCCTCTTTAA 1779
QY 1292 TCTATGAATCTAAGGCTCGGTGTGTGTAACA-ACACACCCCTAAGTGTCTCTTCAAT 1350
Db 1780 TCTATGAATCTAAGGCTCGGTGTGTGTAACAACACACCCCTAAGTGTCTCTTCAAT 1839
QY 1351 ACTCAGCATTTACATTTGTTGAGGCAAAATCAAGCTTTCTCAATCAATTAACAATC 1410
Db 1840 ACTCAGCATTTACATTTGTTGAGGCAAAATCAAGCTTTCTCAATCAATTAACAATC 1899
QY 1411 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCACTGAGTGTGTGTGTAACATGAA 1470
Db 1900 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCACTGAGTGTGTGTGTAACATGAA 1959
QY 1471 GGTAGATCTGAACCCAGGAGTGTCAACAGCTGTCTCACTCCCACTTGGGACTGA 1530
Db 1960 GGTAGATCTGAACCCAGGAGTGTCAACAGCTGTCTCACTCCCACTTGGGACTGA 2019
QY 1531 GAGATATTTCCCTCATTTACTCTCTTAAGGCTATGACCCCTCCACAGCTTCCAGC 1590
Db 2020 GAGATATTTCCCTCATTTACTCTCTTAAGGCTATGACCCCTCCACAGCTTCCAGC 2079
QY 1591 TGGGGAGTGGGGGAGTCAATAGAAAGCCCATCTCCATCTGGATAGGAGACTTCC 1650
Db 2080 TGGGGAGTGGGGGAGTCAATAGAAAGCCCATCTCCATCTGGATAGGAGACTTCC 2139
QY 1651 ATCAGCTTTAACCTTGGGAAATGCTGCTGCTGCCAGTGAATCTTGGTTGCTTCCAC 1710
Db 2140 ATCAGCTTTAACCTTGGGAAATGCTGCTGCTGCCAGTGAATCTTGGTTGCTTCCAC 2199
QY 1711 ATACGAAGCAAGGTGAGGGGAGAGGTGGGCTCAGTTAGAGAGGCTCCCAAGGCAAG 1770
Db 2200 ATACGAAGCAAGGTGAGGGGAGAGGTGGGCTCAGTTAGAGAGGCTCCCAAGGCAAG 2259
QY 1771 TCAGCTCTCTCCCTCATCTCTCTGTGTCAGTGTGCTTAGGTGGGCTCTCACTCCA 1830
Db 2260 TCAGCTCTCTCCCTCATCTCTCTGTGTCAGTGTGCTTAGGTGGGCTCTCACTCCA 2319
QY 1831 CCACCTCTGGG--CCCTTGGGGAGAGCTGGGAGGGGCGCTGGGAGACCTGAAGCTGG 1889
Db 2320 CCACCTCTGGGCGCTTGGGGAGAGCTGGGAGGGGCGCTGGGAGACCTGAAGCTGG 2379
QY 1890 AACCTGTATACAAATTAAGACAGCTCTCAC 1920
Db 2380 AACCTGTATACAAATTAAGACAGCTCTCAC 2410
```

RESULT 13
AAK94832
ID AAK94832 standard; cDNA; 2701 BP.
XX
XX AAK94832;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human full-length cDNA, SEQ ID NO: 3983.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
XX Homo sapiens.

QY	39	CCCAACCAAGCCCGGTGGGGACGGCAATCAATAAGCCGGAGAGCCCTATGAGTTGC	98
Db	833	CACAGCCCGCCCGGTGGGGACGGCAATCAATAAGCCCTATGAGAGCCCTATGAGTTGC	892
QY	99	GGTGAAGCTGAGAGACTTCTCACTGAAGACATGTTGTACCACTCTCCAAACAAACAA	158
Db	893	GGTGAAGCTGAGAGACTTCTCACTGAAGACATGTTGTACCACTCTCCAAACAAACAA	949
QY	159	CATCGAGGTGCGGGGCTGAGAAAGTTGGCGGCTTAACGGACACTGTACATGAACAACCTT	218
Db	950	CATCGAGGTGCGGGGCTGAGAAAGTTGGCGGCTTAACGGACACTGTACATGAACAACCTT	1007
QY	219	ACGACTACACAAGTCCAGACTGCGCGAGACGTGACCCCGAGCGGCGGAGACCTCGG	278
Db	1008	-----CACAAGTGCCAG-CTGCCGAGAGACGTGACCC--GACGTGGGTGACCTCGG	1056
QY	279	CTACTGTGGGAGGAGCAGACAGCTCTACTAGTTCGGGACAGCGTCAACCCGCATACAGAAC	338
Db	1057	---CTCTGGGGAGAGACGGCAGCGCTCTACTAGCGGACAGCGCTCAACCGCATACAGAAC	1113
QY	339	ACGTCACAGACACTTTCGGAGCGAGAGATCAAAATCTGAGTGCCTCCCTTCCTTCC	398
Db	1114	ACGTCACAGACACTTTCGGAGCGAGAGATCAAAATCTGAGTGCCTTCCTTCCTTCC	1173
QY	399	TGTCCCCCGCCCAACGCTGCACAGCAAGGCTTCGTCAACCCATTACAAACAGCTCCAG	458
Db	1174	TGTCCCCCGCCCAACGCTGCACAGCAAGGCTTCGTCAACCCATTACAAACAGCTCCAG	1233
QY	459	ACATCTCAAGCCAGTTTCTAGCCCCCAACGACCCCAAGCCAGGTGACACTCTCCCA	518
Db	1234	ACATCTCAAGCCAGTTTCTAGCCCCCAACGACCCCAAGCCAGGTGACACTCTCCCA	1293

OY	519	AACTAGGGGCGCTCCACTCTATTCACAGGGGACAGGACCTCCCTGAGCCATGATGATG	578
Db	1234	AACTHGGGCCCTCCACTCTATTCACAGGGGACAGGACCTCCCTGAGCCATGATGATG	1355
OY	579	CCCAATTTCAGATTTGGCTTCCTCGTCACTTAATCCAGAGTACAGGGGCTGGGGTCAGGCA	638
Db	1354	CCCAATTTCAGATTTGGCTTCCTCGTCACTTAATCCAGAGTACAGGGGCTGGGGTCAGGCA	1413
OY	639	AGGAAGATCTTAAAGAACCCACTGTGGGCTCAGGGGAAATGGGACAGCAGAGCAATATGGGCA	698
Db	1414	AGGAAGATCTTAAAGAACCCACTGTGGGCTCAGGGGAAATGGGACAGCAGAGCAATATGGGCA	1477
OY	699	AGCTTCGACAGACAGACAGACAAACCTCTGTATCTATGAAGTCTTCGACGGGCAAG	758
Db	1474	AGCTTCGACAGACAGACAGACAAACCTCTGTATCTATGAAGTCTTCGACGGGCAAG	1533
OY	759	GGGACACAGGAGACTGGAAACCTCTTGGCCAGAGGGGAGTGGGAGACAGAGGAAAGTCA	818
Db	1534	GGGACACAGGAGACTGGAAACCTCTTGGCCAGAGGGGAGTGGGAGACAGAGGAAAGTCA	1599
OY	819	CAGGCAAGGGTGCCTATCTAAAGTGAGAACTAATGGCCCGAGGGCTCAGCAAGCCCAAGG	878
Db	1594	CAGGCAAGGGTGCCTATCTAAAGTGAGAACTAATGGCCCGAGGGCTCAGCAAGCCCAAGG	1653
OY	879	AGACAGCCGTACAGGTAAACTTCCCTCTACACGCTCCAGGCCCAAGCCAGCAGCAGAG	938
Db	1654	AGACAGCCGTACAGGTAAACTTCCCTCTACACGCTCCAGGCCCAAGCCAGCAGCAGAG	1713
OY	939	GCTGCTGCTGCCACCCCGTGGCCCAAGCCAGCTGGCTGTGCCAGGGCAGAGCCATGGCACA	998
Db	1714	GCTGCTGCTGCCACCCCGTGGCCCAAGCCAGCTGGCTGTGCCAGGGCAGAGCCATGGCACA	1773
OY	999	TCTGTATATAGATGGGGTTTTCCTCAATACAGCTGTTCTGTAAAAAATGACATGAACCTCC	1058
Db	1774	TCTGTATATAGATGGGGTTTTCCTCAATACAGCTGTTCTGTAAAAAATGACATGAACCTCC	1833
OY	1059	TGCGGTCTCTGGCTGCTGGGGGCTCCAGGCACAGCCAAATGGGGGTTGGGGGTGGGGCTG	1118
Db	1834	TGCGGTCTCTGGCTGCTGGGGGCTCCAGGCACAGCCAAATGGGGGTTGGGGGTGGGGCTG	1893
OY	1119	GTCCTTCTCCCTCCCAAGGCTGTGTCATGTGGGGCTGCTCCATGTAGACAGAGATCAC	1178
Db	1894	GTCCTTCTCCCTCCCAAGGCTGTGTCATGTGGGGCTGCTCCATGTAGACAGAGATCAC	1953
OY	1179	TAAACAGATAGGAAGCCAGGSCATAGATGGGGCTTTGGGTCTTGAGGTTGACCCCAAGC	1238
Db	1954	TAAACAGATAGGAAGCCAGGSCATAGATGGGGCTTTGGGTCTTGAGGTTGACCCCAAGC	2013
OY	1239	TTCTTGGACACTTCCCTCCGCGGAGTCAAGTCCATTCATCCCTCTTAAATCTAATG	1297
Db	2014	TTCTTGGACACTTCCCTCCGCGGAGTCAAGTCCATTCATCCCTCTTAAATCTAATG	2073
OY	1298	AATCTAATAGGCTCGGTGTGTGTAAACAACACCCCTATGGTGTCTTCAAAATATCTAG	1356
Db	2074	AATCTAATAGGCTCGGTGTGTGTAAACAACACCCCTATGGTGTCTTCAAAATATCTAG	2133
OY	1357	CATTACATTTGTGAAGGCCAAATTCAGACTTCTCAAAATCAATTAATCAATCTCATTT	1416
Db	2134	CATTACATTTGTGAAGGCCAAATTCAGACTTCTCAAAATCAATTAATCAATCTCATTT	2193
OY	1417	TTCTATTAAAGGGGAAACATCCCGAGGCACTGAAGTCTGTGCTTTGTCACTGAAGGTTAG	1476
Db	2194	TTCTATTAAAGGGGAAACATCCCGAGGCACTGAAGTCTGTGCTTTGTCACTGAAGGTTAG	2253
OY	1477	ATCGAAGCCAGGGGTGTCAACNGTGTCTCAATCCCAACCTCTGGGCACTGAGAGATA	1536
Db	2254	ATCGAAGCCAGGGGTGTCAACNGTGTCTCAATCCCAACCTCTGGGCACTGAGAGATA	2313
OY	1537	TTTCCCTCAATTTCACTCTCTCTAAGGATAGCAACCTCCCAAGTCTTCAAGTGGGGG	1596
Db	2314	TTTCCCTCAATTTCACTCTCTCTAAGGATAGCAACCTCCCAAGTCTTCAAGTGGGGG	2373

KM	Human; diagnostic; drug screening; forensics; gene mapping;
KM	biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KM	neurodegenerative diseases; anemia; platelet disorder; wound; burns;
KM	ulcers; osteoporosis; autoimmune disease; cancer;
KM	molecular weight marker; food supplement; antiparkinsonian; noctropic;
KM	neuroprotective; antiandemic; anticoagulant; thrombolytic; vulnery;
KM	antifiber; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KM	gene therapy; chromosome 1p36.11-36.2; gene; ss.
XX	
OS	Homo sapiens.
PN	WO2003029271-A2.
XX	
PD	10-APR-2003.
XX	
PF	24-SEP-2002; 2002WO-US030474.
XX	
PR	24-SEP-2001; 2001US-0324631P.
XX	
PA	(HYSEQ-) HYSEQ INC.
PI	Zhang TY, Ghosh M, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T,
P1	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G,
P1	Haley-Vicente D, Drmanac RT;
XX	
DR	WPI; 2003-371981/35.
XX	
P-P	P-PSDB; ADC030955.
XX	
PT	New polynucleotide and polypeptide useful for diagnosing, preventing or
PT	treating conditions such as neurodegenerative diseases, anemias, platelet
PT	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT	cancer.
XX	
PS	Claim 1; SEQ ID NO 66; 1185bp; English.
CC	The invention relates to 971 novel human cDNA sequences (ADC29919-
CC	ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC	invention also relates to nucleic acid sequences over 99% identical with
CC	the novel human cDNAs. The invention additionally encompasses expression
CC	vectors and host cells comprising a nucleic acid of the invention; the
CC	recombinant production of a polypeptide of the invention; an antibody
CC	against a polypeptide of the invention; a method of detecting
CC	polynucleotides or polypeptides of the invention; and methods of
CC	identifying a compound which binds to a polypeptide of the invention. The
CC	invention further discloses methods of preventing, treating or
CC	ameliorating a medical condition; kits comprising polynucleotide probes
CC	and/or monoclonal antibodies for carrying out the methods of the
CC	invention; methods for the identification of compounds that modulate the
CC	expression or activity of the polynucleotide and/or polypeptide; and 767
CC	coding sequences corresponding to the cDNA sequences of the invention
CC	(ADC31861-ADC32627) and the polypeptides encoded by the coding (ADC32628
CC	-ADC33394). The nucleic acids and polypeptides of the invention are
CC	useful in diagnostics, drug screening, forensics, gene mapping, in the
CC	identification of mutations responsible for genetic disorders or other
CC	traits, for assessing biodiversity, and in producing many other types of
CC	data and products dependent on DNA and amino acid sequences. They are
CC	also used for treating diseases such as Parkinson's disease, Alzheimer's
CC	disease and other neurodegenerative diseases, anemia, platelet
CC	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC	cancer. The nucleic acids may also be used as hybridisation probes or
CC	primers, and in the recombinant production of a protein. The polypeptides
CC	are also useful in generating antibodies, as molecular weight markers,
CC	and as food supplements. The present sequence represents a specifically
CC	claimed human cDNA sequence of the invention. Note: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 2783 BP; 586 A; 880 C; 781 G; 536 T; 0 U; 0 Other;
Query Match	82.9%; Score 1648.2; DB 10; Length 2783;
Best Local Similarity	95.3%; Pred. No. 0;
Matches 1812; Conservative	0; Mismatches 60; Indels 29; Gaps 10;

QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGGCAATCAAGACCTTAGAGACCGCTATGA 92
Db 894 GGCCCTTCCAGCCCGCCCGGTGGGAGGCAATCAAGACCTTAGAGACCGCTATGA 953
QY 93 GTTGGCGGTGAGCGTGAAGACTTCTCACTGAGAGACATCAATGTACACCTCCCAACA 152
Db 954 GTTGGCGGTGAGCGTGAAGACTTCTCACTGAGAGACATCAATGTACACCTCCCAACA 1010
QY 153 AACCAACATCGAGGTGCGGAGCTGAGAAAGCTGCGGCTGAAACGACACTGTACATGAA 212
Db 1011 CAACCAATCGAGGTGCGGAGCTGAGAAAGCTGCGGCTGAAACGACACTGTACATGAA 1068
QY 213 CACCTTACGACTACACAAGTGTCCAGACTGCGGAGAGAGTGTGAGACCCGAGCGTGGTGA 272
Db 1069 TTGCGT-----CACAAGTGCAG-CTGCGGAGAGAGTGTGAGACCC--GACGTGGTGA 1117
QY 273 CCTGCGCTACTGTGCGGAGAGAGAGACCTCACTAGTCCGCGGAGCGCTACCGGATA 332
Db 1118 CTTGCGGAG 1174
QY 333 CAGAACACGTCCAGAGACTTCCGAGAGAGATCAAAATCTGATGCTCTCCCTTCCC 392
Db 1175 CAGAACACGTCCAGAGACTTCCGAGAGAGATCAAAATCTGATGCTCTCCCTTCCC 1234
QY 393 TTTCCCTGTCCCGCCCGCCAGCGCTGCGAGAAAGCGCTGCTAACCCCATTAAGAGC 452
Db 1235 TTTCCCTGTCCCGCCCGCCAGCGCTGCGAGAAAGCGCTGCTAACCCCATTAAGAGC 1294
QY 453 TCCAGACATCTAGCCAGGTTTACGCCCCAGAGACCCAGACCCCGAGTGGACCATC 512
Db 1295 TCCAGACATCTAGCCAGGTTTACGCCCCAGAGACCCAGACCCCGAGTGGACCATC 1354
QY 513 CTCCCAACTAGGCGCTCCCACTCTATCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 572
Db 1355 CTCCCAACTAGGCGCTCCCACTCTATCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1414
QY 573 ATGATGCCCAATTTCAATTTGGGCTCCGCTCACTTAATCCAGAGTACAGGGGCTGGGGT 632
Db 1415 ATGATGCCCAATTTCAATTTGGGCTCCGCTCACTTAATCCAGAGTACAGGGGCTGGGGT 1474
QY 633 CAGGAGAGAGATTTAAAGAACCTCTGTGGGTGAGGGGATGAGGACAGAGAGACATA 692
Db 1475 CAGGAGAGAGATTTAAAGAACCTCTGTGGGTGAGGGGATGAGGACAGAGAGACATA 1534
QY 693 TGGGCAAGCTCTGACAGACAGACAGACAGACAAACCTCTGATCTATGAATCTCTGAG 752
Db 1535 TGGGCAAGCTCTGACAGACAGACAGACAGACAAACCTCTGATCTATGAATCTCTGAG 1594
QY 753 GGCAGAGGAGACCAAGGAGCTTGAACCTCTTGGCCAGAGGAGAGAGAGAGAGAGAGAG 812
Db 1595 GGCAGAGGAGACCAAGGAGCTTGAACCTCTTGGCCAGAGGAGAGAGAGAGAGAGAGAG 1654
QY 813 AGGTACACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 872
Db 1655 AGGTACACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1714
QY 873 AAGAGAGAGACAGCGGTGAACGTAACCTTCCCTTACAGGCTCCAGGCTCCAGCGCAGC 932
Db 1715 AAGAGAGAGACAGCGGTGAACGTAACCTTCCCTTACAGGCTCCAGGCTCCAGCGCAGC 1774
QY 933 GAGCAGGCTGCTGCTCCACCCCGTGTCCCGCAGCGAGCTGTGTGTGCGGAGAGCAT 992
Db 1775 GAGCAGGCTGCTGCTCCACCCCGTGTCCCGCAGCGAGCTGTGTGTGCGGAGAGCAT 1834
QY 993 GGCACATCTGTATAGATGAGGAGTTTTCACATACAGCTGGTCTGTGAAAATGACATGA 1052
Db 1835 GGCACATCTGTATAGATGAGGAGTTTTCACATACAGCTGGTCTGTGAAAATGACATGA 1894
QY 1053 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1112
Db 1895 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1954

QY 1113 GGGCTGTCTCTTCT 1172
Db 1955 GGGCTGTCTCTTCT 2014
QY 1173 ATCAGCTTACAGAGATGAG 1232
Db 2015 ATCAGCTTACAGAGATGAG 2074
QY 1233 CCCAGCTTCTTGTGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTA 1291
Db 2075 CCCAGCTTCTTGTGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTA 2134
QY 1292 TCTATGATCTATAGGCTGAGT 1350
Db 2135 TCTATGATCTATAGGCTGAGT 2194
QY 1351 ACTAGCAATTAACCAATTTGAGAGGCAAAATTCAGAGCTTCTCAATCAATCAATCAATC 1410
Db 2195 ACTAGCAATTAACCAATTTGAGAGGCAAAATTCAGAGCTTCTCAATCAATCAATCAATC 2254
QY 1411 TCCATTTTCAATTAACGAGGAGAAATCCCGAGCACTGAGTGTGTGTGTGTGTGTGTGTGT 1470
Db 2255 TCCATTTTCAATTAACGAGGAGAAATCCCGAGCACTGAGTGTGTGTGTGTGTGTGTGTGT 2314
QY 1471 GGTAGATCTGAACCCAGGAGTGTCAACNGCTCTCAATC-----CCGACCTGTGG 1523
Db 2315 GGTAGATCTGAACCCAGGAGTGTCAACNGCTCTCAATC-----CCGACCTGTGG 2374
QY 1524 GCACTAGAGAGATTTTCCCTCTCAATTTACTCTCTTAAGGCTATGCAACCCCTCCAGCTTC 1583
Db 2375 GCACTAGAGAGATTTTCCCTCTCAATTTACTCTCTTAAGGCTATGCAACCCCTCCAGCTTC 2434
QY 1584 TTTCCAGCTGAGGAGTGGGAGGAGTCAAGAGAAAGCCCATCTCCATCTGGATGAGG 1643
Db 2435 TTTCCAGCTGAGGAGTGGGAGGAGTCAAGAGAAAGCCCATCTCCATCTGGATGAGG 2494
QY 1644 ACCTTCATCAGCTTTAACCCTGGAGAAATGCTCTGCTGCTCCCGCAGTACTCTTGGTTGGT 1703
Db 2495 ACCTTCATCAGCTTTAACCCTGGAGAAATGCTCTGCTGCTCCCGCAGTACTCTTGGTTGGT 2554
QY 1704 CTCCCAATTAAG 1763
Db 2555 CTCCCAATTAAG 2614
QY 1764 GGGCAAGTCAAGCTC 1823
Db 2615 GGGCAAGTCAAGCTC 2674
QY 1824 ACTCCCAACACTCTGGG--CCCTTGGGGAGAGAGCTGGGAGAGGGGCGGTGGAGAGCCCTG 1882
Db 2675 ACTCCCAACACTCTGGG--CCCTTGGGGAGAGAGCTGGGAGAGGGGCGGTGGAGAGCCCTG 2734
QY 1883 ACCGTGAACCTGTATACAAATTAAGAGACGTCTCAGAGA 1923
Db 2735 ACCGTGAACCTGTATACAAATTAAGAGACGTCTCAGAGA 2775

Search completed: October 24, 2004, 11:30:45
Job time : 989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2004, 11:14:41 (Search time 970 Seconds
(without alignments)
10496.115 Million cell updates/sec

Title: US-09-977-418-7
Perfect score: 1987
Sequence: 1 tgaatgaagcctgcctatg.....cnnnctacttgagagaa 1987

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3407233 seqs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6814465

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
3: /cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
6: /cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgnt2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgnt2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
13: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
18: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
19: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
20: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
21: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	99.6	1987	10	US-09-977-418-7
2	1979	99.6	1987	10	US-09-977-418-7
3	1979	99.6	1987	10	US-09-977-418-7
4	1979	99.6	1987	10	US-09-977-418-7
5	1979	99.6	1987	10	US-09-977-418-7
6	1979	99.6	1987	10	US-09-977-418-7
7	1979	99.6	1987	10	US-09-977-418-7
8	1979	99.6	1987	10	US-09-977-418-7
9	1979	99.6	1987	10	US-09-977-418-7
10	1979	99.6	1987	10	US-09-977-418-7
11	1979	99.6	1987	10	US-09-977-418-7
12	1979	99.6	1987	10	US-09-977-418-7

13	399.4	20.1	486	10	US-09-918-995-3808	Sequence 3808, Ap
14	375	18.9	400	10	US-09-918-995-4147	Sequence 4147, Ap
15	375	18.9	458	9	US-09-864-761-2033	Sequence 2033, Ap
16	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
17	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
18	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
19	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
20	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
21	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
22	171.4	8.6	719	15	US-10-037-770-669	Sequence 669, Ap
23	154	7.8	472	10	US-09-918-995-5122	Sequence 5122, Ap
24	147	7.4	522	15	US-10-029-186-1250	Sequence 1250, A
25	105.6	5.3	505	10	US-09-918-995-3046	Sequence 3046, Ap
26	105	5.3	134	10	US-09-918-995-1166	Sequence 1166, Ap
27	105	5.3	411	10	US-09-918-995-1761	Sequence 1761, Ap
28	105	5.3	480	10	US-09-918-995-4429	Sequence 4429, Ap
29	105	5.3	482	10	US-09-918-995-5152	Sequence 5152, Ap
30	104	5.2	473	10	US-09-918-995-5122	Sequence 5122, Ap
31	100.6	5.1	462	10	US-09-918-995-5501	Sequence 5501, Ap
32	100.6	5.1	464	10	US-09-918-995-4196	Sequence 4196, Ap
33	85	4.3	196	15	US-10-029-186-1250	Sequence 1250, A
34	75	3.8	192	9	US-09-864-761-1878	Sequence 1878, A
35	60	3.0	60	10	US-09-908-975-1154	Sequence 1154, A
36	60	3.0	60	10	US-09-908-975-1175	Sequence 1175, A
37	60	3.0	60	10	US-09-908-975-1189	Sequence 1189, A
38	60	3.0	60	10	US-09-908-975-1193	Sequence 1193, A
39	54.4	2.7	65	10	US-09-908-975-3947	Sequence 3947, A
40	43	2.2	906	15	US-10-369-493-42907	Sequence 42907, A
41	42.2	2.1	630	17	US-10-437-963-18774	Sequence 18774, A
42	41.8	2.1	674	16	US-10-424-599-69089	Sequence 69089, A
43	41.6	2.1	1049	14	US-10-123-155-358	Sequence 358, Ap
44	41.6	2.1	1049	15	US-10-146-731-358	Sequence 358, Ap
45	41.6	2.1	1049	15	US-10-140-472-358	Sequence 358, Ap

ALIGNMENTS

RESULT 1
US-09-977-418-7
Sequence 7, Application US/09977418
Publication No. US20030027158A1
GENERAL INFORMATION:
APPLICANT: Shinketsu et al
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded t
FILE REFERENCE: 15966-552
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)..(1446)
NAME/KEY: Variation
LOCATION: (1)..(1981)
OTHER INFORMATION: where n can be any nucleotide
US-09-977-418-7
Query Match 99.6%; Score 1979; DB 10; Length 1987;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAGGCTGTCCATGCTGTGCAAGAGCCATCCCAAGAGCCCGGTGGCA 60
Db 1 TGAATGAGGCTGTCCATGCTGTGCAAGAGCCATCCCAAGAGCCCGGTGGCA 60

QY 61 GGCAACTCAAGACCTTAGAGAGCGCTTAGAGTTTGGGTGAGCTGAGAGACTTTCA 120
Db 61 GGCAACTCAAGACCTTAGAGAGCGCTTAGAGTTTGGGTGAGCTGAGAGACTTTCA 120

QY 121 CTTGAAGACATCATTTGTACACCTCCAAACAAACAAACATCGAGGTGGGGCTGAGA 180
Db 121 CTTGAAGACATCATTTGTACACCTCCAAACAAACAAACATCGAGGTGGGGCTGAGA 180

QY 181 AGCTGGCGCTGAACGGAACCTGTACATGAAACACTTACGACTACAAAGTGCAGCT 240
Db 181 AGCTGGCGCTGAACGGAACCTGTACATGAAACACTTACGACTACAAAGTGCAGCT 240

QY 241 GCCGAGAGAGCTGAGACCCCGAGGTGAGCTGAGCTGAGCTGAGGAGAGAGAGAG 300
Db 241 GCCGAGAGAGCTGAGACCCCGAGGTGAGCTGAGCTGAGCTGAGGAGAGAGAGAG 300

QY 301 CTTCACTAGTCCGGGGAACGGGTACCCGGAACAGAAACAGTCCAGACAGCTTCGAGC 360
Db 301 CTTCACTAGTCCGGGGAACGGGTACCCGGAACAGAAACAGTCCAGACAGCTTCGAGC 360

QY 361 GGAATCAAAATGTGAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
Db 361 GGAATCAAAATGTGAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420

QY 421 CAGCAAAAGCTCGCTAACCCATTACAGAGCTCCAGACATCTAGCCAGTTCAGC 480
Db 421 CAGCAAAAGCTCGCTAACCCATTACAGAGCTCCAGACATCTAGCCAGTTCAGC 480

QY 481 CCCCAGACCCCGAGACCCCGAGTGAACATCTCTCCAAACCTAGAGCTTCACTTATC 540
Db 481 CCCCAGACCCCGAGACCCCGAGTGAACATCTCTCCAAACCTAGAGCTTCACTTATC 540

QY 541 CAGGAGAGGCAAGGAGCTCTGAGCTGACATGATGCGCAGATTTGAGATTGGCTC 600
Db 541 CAGGAGAGGCAAGGAGCTCTGAGCTGACATGATGCGCAGATTTGAGATTGGCTC 600

QY 601 CGTCACTTATTCAGAGTACAGGAGGCTGGGCTCAGGAAAGAACTTAAAGAACCACT 660
Db 601 CGTCACTTATTCAGAGTACAGGAGGCTGGGCTCAGGAAAGAACTTAAAGAACCACT 660

QY 661 GTGGGTCAAGGGAATGGAGCAGACGACATATGGGCAAGCTCTGACAGGACAGAGACA 720
Db 661 GTGGGTCAAGGGAATGGAGCAGACGACATATGGGCAAGCTCTGACAGGACAGAGACA 720

QY 721 GACAAACCTCTGATCTATGAAAGTCTTGCAGGGAAGGGAACCAAGGACCTGGAACCT 780
Db 721 GACAAACCTCTGATCTATGAAAGTCTTGCAGGGAAGGGAACCAAGGACCTGGAACCT 780

QY 781 CTTGGCAAGGGGAGTGGAGAGACAGAGGGAAGGTCAAGGCAAGGCTGCTATCTAG 840
Db 781 CTTGGCAAGGGGAGTGGAGAGACAGAGGGAAGGTCAAGGCAAGGCTGCTATCTAG 840

QY 841 TGAATCTAATTGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGTAACTT 900
Db 841 TGAATCTAATTGCCGAGGGCTCAGCAAGGCCAAGAGAGACAGCCGTGACGTAACTT 900

QY 901 CCCCCTTACAGCTTCCAGGCCCAAGCCGAGGAGAGAGGCTGCTGCCACCCGCTGCC 960
Db 901 CCCCCTTACAGCTTCCAGGCCCAAGCCGAGGAGAGAGGCTGCTGCCACCCGCTGCC 960

QY 961 CCAGCAGCTGCTGTGCAAGGGGAGAGCATGCCACATCTGATATGATGGGTTT 1020
Db 961 CCAGCAGCTGCTGTGCAAGGGGAGAGCATGCCACATCTGATATGATGGGTTT 1020

QY 1021 CCAATACAGCTGCTGTGAAAAACTGCATGAATCTCTGCGCTTCTGCGCTGCGG 1080
Db 1021 CCAATACAGCTGCTGTGAAAAACTGCATGAATCTCTGCGCTTCTGCGCTGCGG 1080

Db 1021 CCAATACAGCTGCTGTGAAAAACTGCATGAATCTCTGCGCTTCTGCGCTGCGG 1080

QY 1081 CTTCCAGGCAAGGCAAGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGG 1140
Db 1081 CTTCCAGGCAAGGCAAGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGG 1140

QY 1141 TGTGTTCTTGGGGGTGCTCTCCATGACAGAGATCACTTAAACAGAGTGAAGCAGGCT 1200
Db 1141 TGTGTTCTTGGGGGTGCTCTCCATGACAGAGATCACTTAAACAGAGTGAAGCAGGCT 1200

QY 1201 ATGATGAGGAGCTTGGGGTCCCTGAGGTTGAGAGCCAGCTTCTGCGACCTTCCCGG 1260
Db 1201 ATGATGAGGAGCTTGGGGTCCCTGAGGTTGAGAGCCAGCTTCTGCGACCTTCCCGG 1260

QY 1261 CAGTACAGCTCCATCCATCCCTCTTATATGAACTATAGGCTGGGTGTGTA 1320
Db 1261 CAGTACAGCTCCATCCATCCCTCTTATATGAACTATAGGCTGGGTGTGTA 1320

QY 1321 ACAACACACCCCTATGCTGCTCCCTCAAAATCACTCAGATTAACATTTGAGGCCAAT 1380
Db 1321 ACAACACACCCCTATGCTGCTCCCTCAAAATCACTCAGATTAACATTTGAGGCCAAT 1380

QY 1381 TCAGAGCTTCTCAAAATCAGATTTACATCTCCATTTTCACTTAAACGGGAAACATCCCG 1440
Db 1381 TCAGAGCTTCTCAAAATCAGATTTACATCTCCATTTTCACTTAAACGGGAAACATCCCG 1440

QY 1441 AGCCACTGAGTGTCTGCTTGTCACTGAGGTTAATCTGAACCCAGGCTGACACG 1500
Db 1441 AGCCACTGAGTGTCTGCTTGTCACTGAGGTTAATCTGAACCCAGGCTGACACG 1500

QY 1501 TGCTCTCACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTCACTCTCTA 1560
Db 1501 TGCTCTCACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCACTTCTCACTCTCTA 1560

QY 1561 GGCATATCACCCCTCCCAAGTCTTCAAGTGGGGAGTGGGGAGTCAATGAGAAAGCC 1620
Db 1561 GGCATATCACCCCTCCCAAGTCTTCAAGTGGGGAGTGGGGAGTCAATGAGAAAGCC 1620

QY 1621 CCAATCTCCATCTGGGATGAGGACCTTCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680
Db 1621 CCAATCTCCATCTGGGATGAGGACCTTCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680

QY 1681 CCCCAGTGACTCTTGGTTTCTCTCCCAATCAAGAGGAGTGAAGGAGTGG 1740
Db 1681 CCCCAGTGACTCTTGGTTTCTCTCCCAATCAAGAGGAGTGAAGGAGTGG 1740

QY 1741 GTCTCAGTTAGCAGGGGTCCCGAGGCAAGTCAAGCTCTCCATGCTCTCTGCTC 1800
Db 1741 GTCTCAGTTAGCAGGGGTCCCGAGGCAAGTCAAGCTCTCCATGCTCTCTGCTC 1800

QY 1801 AGTGTGCTTAAAGGTGAGCTCTTCACTCCACCACTCTGAGGCTTGGGGAGGAGCTGGGG 1860
Db 1801 AGTGTGCTTAAAGGTGAGCTCTTCACTCCACCACTCTGAGGCTTGGGGAGGAGCTGGGG 1860

QY 1861 AGGGGCTCGTGGAGAGCCCTGACGCTGGAACCTGTATACAAATTAAGAGACAGTCTCAC 1920
Db 1861 AGGGGCTCGTGGAGAGCCCTGACGCTGGAACCTGTGTATCAAAATTAAGAGACAGTCTCAC 1920

QY 1921 AGACNTCTGAGGCGCCCTGCGGAGGAGTCTCAAACTTGGGCAAGGCGGNNNCTTACTTG 1980
Db 1921 AGACNTCTGAGGCGCCCTGCGGAGGAGTCTCAAACTTGGGCAAGGCGGNNNCTTACTTG 1980

QY 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 2
US-09-977-033A-7
; Sequence 7, Application US/0977033A
; Publication No. US20030082554A1
; GENERAL INFORMATION:
; APPLICANT: Shimkels, Richard A

APPLICANT: Fernandes, Elma
 APPLICANT: Herman, John
 TITLE OF INVENTION: No. US20030082554A1e1 nucleic acid sequences encoding human KIAA0
 TITLE OF INVENTION: protein-like and human protein PRO228-like
 TITLE OF INVENTION: polypeptides.
 FILE REFERENCE: 15966-552 CON-824
 CURRENT APPLICATION NUMBER: US/09/977,033A
 PRIOR FILING DATE: 2001-10-15
 PRIOR APPLICATION NUMBER: 60/137,322
 PRIOR FILING DATE: 1999-06-03
 PRIOR APPLICATION NUMBER: 60/189,810
 PRIOR FILING DATE: 2000-03-16
 PRIOR APPLICATION NUMBER: 60/191,158
 PRIOR FILING DATE: 2000-03-22
 PRIOR APPLICATION NUMBER: 60/193,086
 PRIOR FILING DATE: 2000-03-30
 PRIOR APPLICATION NUMBER: 60/201,388
 PRIOR FILING DATE: 2000-05-03
 PRIOR APPLICATION NUMBER: 09/584,411
 PRIOR FILING DATE: 2000-05-31
 NUMBER OF SEQ ID NOS: 92
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 7
 LENGTH: 1987
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (991)..(1446)
 NAME/KEY: variation
 LOCATION: (1148)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: (1498)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1925)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1943)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1969)..(1972)
 OTHER INFORMATION: wherein n is an a or t or c or g
 US-09-977-033A-7
 Query Match 99.6%; Score 1979; DB 10; Length 1987;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 301 CCTCACTAGTCCGGGACGCGTCAACCGCATACAGAACGCTCCAGACGACTTCGGAC 360
 Qy 361 GGAGATCAAAATCTGAGTGCCTTCCTCCCTTTCCTTCCTCCCGCCCGCCAGCCTGC 420
 Db 361 GGAGATCAAAATCTGAGTGCCTTCCTCCCTTTCCTTCCTCCCGCCCGCCAGCCTGC 420
 Qy 421 CAGCAAGCCTTGCTTAACCCCATTAACAACAGCTCCAGAGACATTCAGGCCAGTCTAGC 480
 Db 421 CAGCAAGCCTTGCTTAACCCCATTAACAACAGCTCCAGAGACATTCAGGCCAGTCTAGC 480
 Qy 481 CCCCAAGACCCAGACCCCAAGTGAACCATCTCCCAATAGAGCCCTCACTATAC 540
 Db 481 CCCCAAGACCCAGACCCCAAGTGAACCATCTCCCAATAGAGCCCTCACTATAC 540
 Qy 541 CAGGCGAGGCCAGGAGCTCCCTGCTGACATATATGCCAGATTTAGATTGGCTTC 600
 Db 541 CAGGCGAGGCCAGGAGCTCCCTGCTGACATATATGCCAGATTTAGATTGGCTTC 600
 Qy 601 CTTCACTTAATCCAGATACAGGGGCTGGGGTCAAGGAGAAATCTAAAGAACCACT 660
 Db 601 CTTCACTTAATCCAGATACAGGGGCTGGGGTCAAGGAGAAATCTAAAGAACCACT 660
 Qy 661 GTGGGTCAAGGAGATGAGACAGAGGAGATCAAGGCAAGGCTCTGACAGACAGACACA 720
 Db 661 GTGGGTCAAGGAGATGAGACAGAGGAGATCAAGGCAAGGCTCTGACAGACAGACACA 720
 Qy 721 GACAAACCTTCTGATCTATGAAATCTCTGACAGGCGCAAGGACCTGGAACCT 780
 Db 721 GACAAACCTTCTGATCTATGAAATCTCTGACAGGCGCAAGGACCTGGAACCT 780
 Qy 781 CTGGCCAAAGGAGATGAGAGAGACAGAGGAGATCAAGGCAAGGCTCTGACAGAC 840
 Db 781 CTGGCCAAAGGAGATGAGAGAGACAGAGGAGATCAAGGCAAGGCTCTGACAGAC 840
 Qy 841 TGGAACTAATGCCCCAGGGCTCAGCAAGGCCAAGAGACAGCCGCTGAACTT 900
 Db 841 TGGAACTAATGCCCCAGGGCTCAGCAAGGCCAAGAGACAGCCGCTGAACTT 900
 Qy 901 CCCCTTACAGGCTTCCAGAGCCCAAGGAGAGAGGCTGCTGCCACCCCGTGGCC 960
 Db 901 CCCCTTACAGGCTTCCAGAGCCCAAGGAGAGAGGCTGCTGCCACCCCGTGGCC 960
 Qy 961 CCAGCAGCTGGCTGTGCCAGGAGAGGCAATGCCATCTGTATATAGTGGGTTTTT 1020
 Db 961 CCAGCAGCTGGCTGTGCCAGGAGAGGCAATGCCATCTGTATATAGTGGGTTTTT 1020
 Qy 1021 CCAATACAGCTGGTGTGTAAGAACTGCAAGAACTCCGCGCTGCGGCTGGGG 1080
 Db 1021 CCAATACAGCTGGTGTGTAAGAACTGCAAGAACTCCGCGCTGCGGCTGGGG 1080
 Qy 1081 CTTCCAGGCAAGGCCAAGTGGGTTGGGGTGGGGTGGGCTCTTCCCTCCACAGGCC 1140
 Db 1081 CTTCCAGGCAAGGCCAAGTGGGTTGGGGTGGGGTGGGCTCTTCCCTCCACAGGCC 1140
 Qy 1141 TGTGTTCTTGGGGCTGTCTCCATGCAAGATCACTTAACAGAAATGGAAGCCAGGCC 1200
 Db 1141 TGTGTTCTTGGGGCTGTCTCCATGCAAGATCACTTAACAGAAATGGAAGCCAGGCC 1200
 Qy 1201 ATGATGGGGCTTGGGCTCTGAGGTGGAACCCAGCTTCTTCCACCTCCCGG 1260
 Db 1201 ATGATGGGGCTTGGGCTCTGAGGTGGAACCCAGCTTCTTCCACCTCCCGG 1260
 Qy 1261 CAGTACCTCTTCATCATCCCTCTTTATATATATATATATATATATATATATATAT 1320
 Db 1261 CAGTACCTCTTCATCATCCCTCTTTATATATATATATATATATATATATATATAT 1320
 Qy 1321 ACAACACACCCCTATGCTTCTTCTTAATATCAATATCAATATCAATATCAATATCA 1380
 Db 1321 ACAACACACCCCTATGCTTCTTCTTAATATCAATATCAATATCAATATCAATATCA 1380
 Qy 1381 TCAGAGCTTTCTCAATATCAATATCAATATCAATATCAATATCAATATCAATATCA 1440
 Db 1381 TCAGAGCTTTCTCAATATCAATATCAATATCAATATCAATATCAATATCAATATCA 1440

QY 601 CGTCACTTAATCCAGAGTACAGGGGCTGGGGCTCGAGGAGAGAGATCTAAGAACCCACT 660
Db 601 CGTCACTTAATCCAGAGTACAGGGGCTGGGGCTCGAGGAGAGAGATCTAAGAACCCACT 660
QY 661 GTGGGTGAGGGGAAATGGGACCAAGAGACATATGGGCAAGCTCTGACAGGACAGACAGACA 720
Db 661 GTGGGTGAGGGGAAATGGGACCAAGAGACATATGGGCAAGCTCTGACAGGACAGACAGACA 720
QY 721 GACAAACCCCTGTATCTATGAAAGTCTGTGACAGGCAAGGGGACCAAGGAGCCTGGAACTT 780
Db 721 GACAAACCCCTGTATCTATGAAAGTCTGTGACAGGCAAGGGGACCAAGGAGCCTGGAACTT 780
QY 781 CTTGGCCCAAGGGAGTGGGAGAGACAGAGGAGAGTCAACAGGCAAGGCTCTATCTAG 840
Db 781 CTTGGCCCAAGGGAGTGGGAGAGACAGAGGAGAGTCAACAGGCAAGGCTCTATCTAG 840
QY 841 TGAACCTAATTTGCCGAGAGGCTCAGCAAGGCCAAGAGAGACAGCCGTGAAGGTAACCTT 900
Db 841 TGAACCTAATTTGCCGAGAGGCTCAGCAAGGCCAAGAGAGACAGCCGTGAAGGTAACCTT 900
QY 901 CCCCTCTACAGAGCTCCAGGCCCAAGCCGACGACAGAGCTGCTGCTCCACCCCTGCTCC 960
Db 901 CCCCTCTACAGAGCTCCAGGCCCAAGCCGACGACAGAGCTGCTGCTCCACCCCTGCTCC 960
QY 961 CCAGCCAGCTGGCTGTGCGAGGGGAGAGCCATGCGACATCTGTATATAGATGGGGTTTTT 1020
Db 961 CCAGCCAGCTGGCTGTGCGAGGGGAGAGCCATGCGACATCTGTATATAGATGGGGTTTTT 1020
QY 1021 CCAATATACAGTGTGCTGTGAAATACTGCATGAAATCTCTGCGCTCTGCGCTGCTG 1080
Db 1021 CCAATATACAGTGTGCTGTGAAATACTGCATGAAATCTCTGCGCTCTGCGCTGCTG 1080
QY 1081 CCTCAGGCAAGGCCAAGTGGGGGTTGGGGGTTGGGGGTTGGTCTCTCTCTCCCAAGGGC 1140
Db 1081 CCTCAGGCAAGGCCAAGTGGGGGTTGGGGGTTGGGGGTTGGTCTCTCTCTCCCAAGGGC 1140
QY 1141 TGTGTTCTGGGGGCTGTCTCCATGACAGACAGATCACTTAACAGAGATGAGAGCAGAGC 1200
Db 1141 TGTGTTCTGGGGGCTGTCTCCATGACAGACAGATCACTTAACAGAGATGAGAGCAGAGC 1200
QY 1201 ATGATGAGGGCTTTGGGCTCTGAGAGTTGAGACCCAGCTTCTTGACCTCTCCCTCG 1260
Db 1201 ATGATGAGGGCTTTGGGCTCTGAGAGTTGAGACCCAGCTTCTTGACCTCTCCCTCG 1260
QY 1261 CAGTACAGCTCCATCCATCCCTCTTATATCTATGATATGATAGGCTGGGTGTGTGTA 1320
Db 1261 CAGTACAGCTCCATCCATCCCTCTTATATCTATGATATGATAGGCTGGGTGTGTGTA 1320
QY 1321 ACAACACACCCCTATCGTTGCTTCAAAATCTCAGACATTAACATTGTTGAGGCCAAT 1380
Db 1321 ACAACACACCCCTATCGTTGCTTCAAAATCTCAGACATTAACATTGTTGAGGCCAAT 1380
QY 1381 TCAAGGCTTCTCAAAATCAAGATTAACATCTCCATTTCAATTAAGGGGAGAACATCCCG 1440
Db 1381 TCAAGGCTTCTCAAAATCAAGATTAACATCTCCATTTCAATTAAGGGGAGAACATCCCG 1440
QY 1441 AGCCACTGAGTGTGCTTGTCTGCACTGAGAGTTAGATCTGACCCAGGCTGTCAACNGC 1500
Db 1441 AGCCACTGAGTGTGCTTGTCTGCACTGAGAGTTAGATCTGACCCAGGCTGTCAACNGC 1500
QY 1501 TGCTCTCAACTCCCACTCTGGGACAGAGAGATTTCCCTTATCTCACTCTCTAA 1560
Db 1501 TGCTCTCAACTCCCACTCTGGGACAGAGAGATTTCCCTTATCTCACTCTCTAA 1560
QY 1561 GGCATGACCCCTCCCACTCTGCTTCAAGCTGGGGAGTGTATGAGAAAGCC 1620
Db 1561 GGCATGACCCCTCCCACTCTGCTTCAAGCTGGGGAGTGTATGAGAAAGCC 1620
QY 1621 CCCATCTCCCATCTGGAGATGAGGACCTTCCATCAGCCTTAACTCTGGGAAATGCTGTG 1680
Db 1621 CCCATCTCCCATCTGGAGATGAGGACCTTCCATCAGCCTTAACTCTGGGAAATGCTGTG 1680

QY 1681 CCCCCAGTACTCTTGGTTTGGTCTCCCATACAGAAAGAGGGTGAAGGGGAGGGTGG 1740
Db 1681 CCCCCAGTACTCTTGGTTTGGTCTCCCATACAGAAAGAGGGTGAAGGGGAGGGTGG 1740
QY 1741 GTCTCAGTTAGAGGGGTTCCCAAGGCAAGTACAGCTCTCTCCATGCTCTGTGTC 1800
Db 1741 GTCTCAGTTAGAGGGGTTCCCAAGGCAAGTACAGCTCTCTCTCCATGCTCTGTGTC 1800
QY 1801 AGTGTGCTTACAGGTTGCTCTCACTCCACCACTCTGGGCTTGGGGAGAGACTGGGG 1860
Db 1801 AGTGTGCTTACAGGTTGCTCTCACTCCACCACTCTGGGCTTGGGGAGAGACTGGGG 1860
QY 1861 AGGGGGCGTGGGAGAGCCCTGACGCTGGAACCTGTATACATAAAGACAGTCTAC 1920
Db 1861 AGGGGGCGTGGGAGAGCCCTGACGCTGGAACCTGTATACATAAAGACAGTCTAC 1920
QY 1921 AGACNTCTGAGGCGCGCTGCGNGAGATTCTCAAACTTAGGGCAGGGCANNCTTACTTG 1980
Db 1921 AGACNTCTGAGGCGCGCTGCGNGAGATTCTCAAACTTAGGGCAGGGCANNCTTACTTG 1980
QY 1981 AGAGAA 1987
Db 1981 AGAGAA 1987

RESULT 4
US-09-977-639A-7
Sequence 7, Application US/09977639A
Publication No. US20030199103A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Heriman, John
TITLE OF INVENTION: No. US20030199103A1: amino acid sequences for human epidermal 9
FILE REFERENCE: 15966-552 CON-S34
CURRENT APPLICATION NUMBER: US/09/977, 639A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991) ..(1446)
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1148) ..(1148)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1498) ..(1498)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1925) ..(1925)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:

NAME/KEY: Misc_Feature
LOCATION: (1943)...(1943)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc_Feature
LOCATION: (1969)...(1972)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Variant
LOCATION: (53)...(53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.
US-09-977-639A-7

Query Match 99.6%; Score 1979; DB 10; Length 1987;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAAGGCTGTCCATCTCTGTCTAAAGAGCCCATCCCAACCAAGCCCGGTGGGA 60
DB 1 TGAATGAAGGCTGTCCATCTCTGTCTAAAGAGCCCATCCCAACCAAGCCCGGTGGGA 60
QY 61 GGCAACTCAAGACCTTAGAGAGCCCTATGAGTTTGGGTGACGTGAGACTTTCTCA 120
DB 61 GGCAACTCAAGACCTTAGAGAGAGCCCTATGAGTTTGGGTGACGTGAGACTTTCTCA 120
QY 121 CTTGAAGACATCATTTGTACACCTCCAAACAAACCAATCGAGGTGGGGCTGAGAA 180
DB 121 CTTGAAGACATCATTTGTACACCTCCAAACAAACCAATCGAGGTGGGGCTGAGAA 180
QY 181 AGCTGGCGCTGAAACCGACATCTATACATGAAACACTTAAGACTACAAAGTCCAGACT 240
DB 181 AGCTGGCGCTGAAACCGACATCTATACATGAAACACTTAAGACTACAAAGTCCAGACT 240
QY 241 GCGGAGAGAGCTGAGACCCCGAGCTGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 300
DB 241 GCGGAGAGAGCTGAGACCCCGAGCTGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 300
QY 301 CCTACTAGTCCCGGACGCGTCAACCGGATACAGAAACGTCACAGACCTTCGAGAC 360
DB 301 CCTACTAGTCCCGGACGCGTCAACCGGATACAGAAACGTCACAGACCTTCGAGAC 360
QY 361 GGAGATCAAAATTTGAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 420
DB 361 GGAGATCAAAATTTGAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 420
QY 421 CAGCAAGGCTCGCTAACCCCATTTACACAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
DB 421 CAGCAAGGCTCGCTAACCCCATTTACACAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
QY 481 CCCACGCAACCCGACCCGAGGTGAGCCATCTCCCAAACTAGGGCCCTCCACTCTATC 540
DB 481 CCCACGCAACCCGACCCGAGGTGAGCCATCTCCCAAACTAGGGCCCTCCACTCTATC 540
QY 541 CAGGAGAGGAGGAGCTCCCTGGCTGACATGATGCGCAGATTTTCAAGTTTGGGCTC 600
DB 541 CAGGAGAGGAGGAGCTCCCTGGCTGACATGATGCGCAGATTTTCAAGTTTGGGCTC 600
QY 601 CGTCACTTATTCAGAGTAAGAGGCTGAGGCTCAGGGAAGAGATCTTAAAGAACCCACT 660
DB 601 CGTCACTTATTCAGAGTAAGAGGCTGAGGCTCAGGGAAGAGATCTTAAAGAACCCACT 660
QY 661 GTGGGTCAGGGGAATGGGACAGAGACATATGGGCAAGCTCTGACAGACAGACAGCA 720
DB 661 GTGGGTCAGGGGAATGGGACAGAGACATATGGGCAAGCTCTGACAGACAGACAGCA 720
QY 721 GACAAACCTCTGATCTATGAGTCTCTGACAGGGGACAGGGGACCTGGAACCT 780
DB 721 GACAAACCTCTGATCTATGAGTCTCTGACAGGGGACAGGGGACCTGGAACCT 780
QY 781 CTTGGGCAAGGGAAGTGGAGAGACAGAGGAAGGTCAAGGCAAGGCTGCTATCAAG 840
DB 781 CTTGGGCAAGGGAAGTGGAGAGACAGAGGAAGGTCAAGGCAAGGCTGCTATCAAG 840

QY 841 TGAATCTATTGCGGAGGCTCAGCAAGGCCAAGAGAGACAGCCGTTGAAGTAACTT 900
DB 841 TGAATCTATTGCGGAGGCTCAGCAAGGCCAAGAGAGACAGCCGTTGAAGTAACTT 900
QY 901 CCCCTCTACAGGCTTCCAGAGCCCAACGCGACAGGACGAGCTGCTGCCACCCGTCGCC 960
DB 901 CCCCTCTACAGGCTTCCAGAGCCCAACGCGACAGGACGAGCTGCTGCCACCCGTCGCC 960
QY 961 CCAGCAGCTGCTGTGCGCAGGAGAGAGCCATGCAATCTGATATATAGATGGGGTTTT 1020
DB 961 CCAGCAGCTGCTGTGCGCAGGAGAGAGCCATGCAATCTGATATATAGATGGGGTTTT 1020
QY 1021 CCAATACAGCTGTGTTCTGAAAAAAGTGAATGAACTCTGCGCTCTGCGCTGTCGGG 1080
DB 1021 CCAATACAGCTGTGTTCTGAAAAAAGTGAATGAACTCTGCGCTCTGCGCTGTCGGG 1080
QY 1081 CCTCAGGCAAGGCAAGTGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1140
DB 1081 CCTCAGGCAAGGCAAGTGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1140
QY 1141 TGTGTTCTGTTGAGGCTGTCTCCATGACAGAGATCACTTAACAGAGATGGAAGCCAGGC 1200
DB 1141 TGTGTTCTGTTGAGGCTGTCTCCATGACAGAGATCACTTAACAGAGATGGAAGCCAGGC 1200
QY 1201 ATGATGGGGCTTGGGGTCTGAGGTTGAGACCCAGCTCTGCAACCTTCCCTCCGG 1260
DB 1201 ATGATGGGGCTTGGGGTCTGAGGTTGAGACCCAGCTCTGCAACCTTCCCTCCGG 1260
QY 1261 CAGTCAAGCTTCAATCCCTCCCTCTTATATCATGAACTTAAGCTGAGTGTGTA 1320
DB 1261 CAGTCAAGCTTCAATCCCTCCCTCTTATATCATGAACTTAAGCTGAGTGTGTA 1320
QY 1321 ACAAACAGCCCTATGCTGTCTTCAAAATCAAGATTAACATTTGTTGAGGCCAAAT 1380
DB 1321 ACAAACAGCCCTATGCTGTCTTCAAAATCAAGATTAACATTTGTTGAGGCCAAAT 1380
QY 1381 TCAAGGCTTCTCAAAATCAAGATTTCAATCTCAATTTCTAATCAAGGGGAAACATCCCG 1440
DB 1381 TCAAGGCTTCTCAAAATCAAGATTTCAATCTCAATTTCTAATCAAGGGGAAACATCCCG 1440
QY 1441 AGCCATGAGTGTCTGTCTTGTCTGACAGGTTAATGTTCAACCAAGGTTCAACGCG 1500
DB 1441 AGCCATGAGTGTCTGTCTTGTCTGACAGGTTAATGTTCAACCAAGGTTCAACGCG 1500
QY 1501 TGCTCTCAATCCCACTCTGAGGACAGTGAAGTATTTCCCTCAATCTCTCTCTA 1560
DB 1501 TGCTCTCAATCCCACTCTGAGGACAGTGAAGTATTTCCCTCAATCTCTCTCTA 1560
QY 1561 GGTATGACACCTTCCCACTCTTCAAGTGGGGATGGGGAGTCAATAGGAAAGCC 1620
DB 1561 GGTATGACACCTTCCCACTCTTCAAGTGGGGATGGGGAGTCAATAGGAAAGCC 1620
QY 1621 CCAATCTCCATCTGGGATGGGACCTTCCATGAGCTTAACTTGGAAATGCTGTCTG 1680
DB 1621 CCAATCTCCATCTGGGATGGGACCTTCCATGAGCTTAACTTGGAAATGCTGTCTG 1680
QY 1681 CCCCAGTGACTCTGTGTTCTGCTCCCATACAGAGAGAGGTTGAGAGGAGGTTG 1740
DB 1681 CCCCAGTGACTCTGTGTTCTGCTCCCATACAGAGAGAGGTTGAGAGGAGGTTG 1740
QY 1741 GTCTCAGTTAGCAGGGGTCCCAAGGCAAGTCAAGCTCTCTCCATGCTCTGTGTC 1800
DB 1741 GTCTCAGTTAGCAGGGGTCCCAAGGCAAGTCAAGCTCTCTCCATGCTCTGTGTC 1800
QY 1801 AGTGTGCTTAAAGGCTGCTCACTCCACACTCTGAGGCTTGGGGGAGAGCTGGGG 1860
DB 1801 AGTGTGCTTAAAGGCTGCTCACTCCACACTCTGAGGCTTGGGGGAGAGCTGGGG 1860
QY 1861 AGGGGGCTGGGGAGGCTCTGACGCTGGAACCTGTATACAAATTAAGAGACAGTCTAC 1920
DB 1861 AGGGGGCTGGGGAGGCTCTGACGCTGGAACCTGTATACAAATTAAGAGACAGTCTAC 1920
QY 1921 AGACNTCTGAGAGGCCCTGCGCAGAGTCTCAAACTTAGGCGAGGCGNNNNCTACTTG 1980

Db 1921 AGACNTCTGGAGGCGCTGCGGAGTTCTCAACTAGGCGGCGNNNNCTTACTTG 1980
 Qy 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987

RESULT 5

US-09-977-819B-7
 : Sequence 7, Application US/09977819B
 : Publication No. US20040002134A1

GENERAL INFORMATION:
 : APPLICANT: Shimkels, Richard A
 : APPLICANT: Hernandez, Elma
 : APPLICANT: Herrman, John
 : APPLICANT: Vernet, Corine
 : TITLE OF INVENTION: No. US20040002134A1el nucleic acid sequences encoding human KIAA
 : TITLE OF INVENTION: Protein-like and human protein PRO228-like
 : FILE REFERENCE: 15966-552 CON-526
 : CURRENT APPLICATION NUMBER: US/09/977,819B
 : PRIOR FILING DATE: 2001-10-15
 : PRIOR APPLICATION NUMBER: 60/137,322
 : PRIOR FILING DATE: 1999-06-03
 : PRIOR APPLICATION NUMBER: 60/189,810
 : PRIOR FILING DATE: 2000-03-16
 : PRIOR APPLICATION NUMBER: 60/191,158
 : PRIOR FILING DATE: 2000-03-22
 : PRIOR APPLICATION NUMBER: 60/193,086
 : PRIOR FILING DATE: 2000-03-30
 : PRIOR APPLICATION NUMBER: 60/201,388
 : PRIOR FILING DATE: 2000-05-03
 : PRIOR APPLICATION NUMBER: 09/584,411
 : PRIOR FILING DATE: 2000-05-31
 : NUMBER OF SEQ ID NOS: 92
 : SOFTWARE: Patent In Ver. 2.1
 : SEQ ID NO 7
 : LENGTH: 1987
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (991)..(1446)
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (1145)..(1145)
 : OTHER INFORMATION: "n" = "a", "c", "t" or "g"
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (1495)..(1495)
 : OTHER INFORMATION: "n" = "a", "c", "t" or "g"
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (1925)..(1925)
 : OTHER INFORMATION: "n" = "a", "c", "t" or "g"
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (1943)..(1943)
 : OTHER INFORMATION: "n" = "a", "c", "t" or "g"
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (1969)..(1972)
 : OTHER INFORMATION: "n" = "a", "c", "t" or "g"
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (533)..(533)
 : OTHER INFORMATION: "xaa" = "ile", "leu", "val" or "phe"
 : US-09-977-819B-7

Query Match 99.6%; Score 1979; DB 11; Length 1987;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1987; Conservative 0; Mismatches 0; Gaps 0;

Qy	1	TGATAGAGAGGCTGTGCTCATGCTCTGTCAAGAGAGCCCATCCCAACCAAGCCCGGTGGCA	60
Db	1	TGATAGAGAGGCTGTGCTCATGCTCTGTCAAGAGAGCCCATCCCAACCAAGCCCGGTGGCA	60
Qy	61	GGCAACATCAAGACCTTAGAGAGCGCTTATGAGTTTGGGTGAGCTGAGACTTCTCA	120
Db	61	GGCAACATCAAGACCTTAGAGAGCGCTTATGAGTTTGGGTGAGCTGAGACTTCTCA	120
Qy	121	CCTGAAGACATCATTTGTCAACCACTTCAAAACCAACCAATGAGGTGGGGCTGAGAA	180
Db	121	CCTGAAGACATCATTTGTCAACCACTTCAAAACCAACCAATGAGGTGGGGCTGAGAA	180
Qy	181	AGCTGGCGGCTGAGCGGACATGTATGATGAACCTTACGACTACAAAGTGCAGACT	240
Db	181	AGCTGGCGGCTGAGCGGACATGTATGATGAACCTTACGACTACAAAGTGCAGACT	240
Qy	241	GCCGAGAGAGGTGAGACCCGAGGCGGTGAGCTGCGTACTGTGCGGAGGAGCGGACAG	300
Db	241	GCCGAGAGAGGTGAGACCCGAGGCGGTGAGCTGCGTACTGTGCGGAGGAGCGGACAG	300
Qy	301	CCTCACTAGTCCGGGACGCGTACCCGCTACAGAACAGTCCAGACAGCTTCGGAC	360
Db	301	CCTCACTAGTCCGGGACGCGTACCCGCTACAGAACAGTCCAGACAGCTTCGGAC	360
Qy	361	GGAGATCAAAATGTGAGTGCCTCTCCCTTCCCTTCCCTGTCGCCCGCCGACAGCTGC	420
Db	361	GGAGATCAAAATGTGAGTGCCTCTCCCTTCCCTTCCCTGTCGCCCGCCGACAGCTGC	420
Qy	421	CAGCAAGCCTGCTACCCCATTTACAGAGCTCCAGACATCTCAGCCAGGTTCTAC	480
Db	421	CAGCAAGCCTGCTACCCCATTTACAGAGCTCCAGACATCTCAGCCAGGTTCTAC	480
Qy	481	CCCCAGCACCCCAAGAGCCGAGTGGACATCTTCCCAACTTAGGGCCTTCACTTAC	540
Db	481	CCCCAGCACCCCAAGAGCCGAGTGGACATCTTCCCAACTTAGGGCCTTCACTTAC	540
Qy	541	CAGGAGAGGAGGAGGAGCTCCCTGCTGACATGATGAGCCAGATTTCAGTTTGACCTC	600
Db	541	CAGGAGAGGAGGAGGAGCTCCCTGCTGACATGATGAGCCAGATTTCAGTTTGACCTC	600
Qy	601	CGTCACTTAATCCAGATACAGGAGGCTGGGATGAGGAGGAGGAGATCTAAAGAACCCACT	660
Db	601	CGTCACTTAATCCAGATACAGGAGGCTGGGATGAGGAGGAGGAGATCTAAAGAACCCACT	660
Qy	661	GTGGGTCAAGGGGAAATGGGACAGAGGACATATGGGCAAGCTCTGAGGACAGACAGACA	720
Db	661	GTGGGTCAAGGGGAAATGGGACAGAGGACATATGGGCAAGCTCTGAGGACAGACAGACA	720
Qy	721	GACAAACCTCTGATCTATGAAGTCTCTGACAGGACAGAGGACCTGGAACCTCT	780
Db	721	GACAAACCTCTGATCTATGAAGTCTCTGACAGGACAGAGGACCTGGAACCTCT	780
Qy	781	CTTGCCCAAGGGAGTGGAGAGACAGAGGAGAGTCAAGGAGAGGAGTCTATTAAG	840
Db	781	CTTGCCCAAGGGAGTGGAGAGACAGAGGAGAGTCAAGGAGAGGAGTCTATTAAG	840
Qy	841	TGGAATTAATTTGGCCAGAGGCTCAGCAAGGCAAGAGAGAGAGCCGTATTAAGT	900
Db	841	TGGAATTAATTTGGCCAGAGGCTCAGCAAGGCAAGAGAGAGAGCCGTATTAAGT	900
Qy	901	CCCTCTTACCAAGCTTCCAAAGCCCAAGGAGAGGAGGCTGCTCCACCCCGTGGCC	960
Db	901	CCCTCTTACCAAGCTTCCAAAGCCCAAGGAGAGGAGGCTGCTCCACCCCGTGGCC	960
Qy	961	CCAGCAGAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1020
Db	961	CCAGCAGAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1020
Qy	1021	CCAATACAGCTGTTGTTGTAAGAAATGATGAATCTCTGCGCTGCGCTGCTGCGG	1080
Db	1021	CCAATACAGCTGTTGTTGTAAGAAATGATGAATCTCTGCGCTGCGCTGCTGCGG	1080

QY 1081 CCTCCAGGCAAGCCAAAGTGGGGTGGGGGCTGGGCTCTTCTCCCTCCCAAGGCC 1140
 DB 1081 CCTCCAGGCAAGCCAAAGTGGGGTGGGGGCTGGGCTCTTCTCCCTCCCAAGGCC 1140
 QY 1141 TGTGTTGTTGGGGTGGTCTCCATGTCAGACAGATCACTTAACAGAGATGAAAGCCAGGC 1200
 DB 1141 TGTGTTGTTGGGGTGGTCTCCATGTCAGACAGATCACTTAACAGAGATGAAAGCCAGGC 1200
 QY 1201 ATGATGAGGGCTTTGGGCTCTGAGGTGGAACCCAGCTTTCTTGCCACTTCCCTCCG 1260
 DB 1201 ATGATGAGGGCTTTGGGCTCTGAGGTGGAACCCAGCTTTCTTGCCACTTCCCTCCG 1260
 QY 1261 CAGTCAAGCTCTCATCATCCCTCTCTTAATCTAATGATCTAATAGGCTCGGTGTGA 1320
 DB 1261 CAGTCAAGCTCTCATCATCCCTCTCTTAATCTAATGATCTAATAGGCTCGGTGTGA 1320
 QY 1321 ACAACACACCCCTATGTTGCTTCTTCAAACTACAGCATTAACCATTTGTTAGGCCAAT 1380
 DB 1321 ACAACACACCCCTATGTTGCTTCTTCAAACTACAGCATTAACCATTTGTTAGGCCAAT 1380
 QY 1381 TCAGAGCTTCTCAAACTAGATTTACATCTCAATTTTCAATTAAGGGGAAATCTCCG 1440
 DB 1381 TCAGAGCTTCTCAAACTAGATTTACATCTCAATTTTCAATTAAGGGGAAATCTCCG 1440
 QY 1441 AGCCACTGAGTGTGCTTGTCTTGTCACTGAAGTTAGATCTGAACCCAGGCTGCAACGC 1500
 DB 1441 AGCCACTGAGTGTGCTTGTCTTGTCACTGAAGTTAGATCTGAACCCAGGCTGCAACGC 1500
 QY 1501 TGTCTTCAACTCTCCCACTCTTGGGCACTGAGAGATTTTCCCTCACTTACCTCTTA 1560
 DB 1501 TGTCTTCAACTCTCCCACTCTTGGGCACTGAGAGATTTTCCCTCACTTACCTCTTA 1560
 QY 1561 GGTATGACACCCCTCCCAAGCTCTTCCAGCTGGGGATGGGGGAGTCACTAGAAAAGCC 1620
 DB 1561 GGTATGACACCCCTCCCAAGCTCTTCCAGCTGGGGATGGGGGAGTCACTAGAAAAGCC 1620
 QY 1621 CCATCTCCCATCTGGGATAGGAGCTTCCATCAGCTTAACCTTGGGAAATGCTGCTG 1680
 DB 1621 CCATCTCCCATCTGGGATAGGAGCTTCCATCAGCTTAACCTTGGGAAATGCTGCTG 1680
 QY 1681 CCCCAGTGACTCTTGGTTGCTCTCCCACTACAGAGAGGAGGAGGAGGAGGAGGAGG 1740
 DB 1681 CCCCAGTGACTCTTGGTTGCTCTCCCACTACAGAGAGGAGGAGGAGGAGGAGGAGG 1740
 QY 1741 GTCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCCCTCCATGCTCTGCTG 1800
 DB 1741 GTCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCCCTCCATGCTCTGCTG 1800
 QY 1801 AGTGTGCTTGAAGGTGGCTCTCACTCCCACTGAGGCTCTTGGGGGAGGAGGAGGAGG 1860
 DB 1801 AGTGTGCTTGAAGGTGGCTCTCACTCCCACTGAGGCTCTTGGGGGAGGAGGAGGAGG 1860
 QY 1861 AGGGGCGGTGGAGAGCCCTGACGTGAACTGTATACAAATTAAGGACAGTCTGAC 1920
 DB 1861 AGGGGCGGTGGAGAGCCCTGACGTGAACTGTATACAAATTAAGGACAGTCTGAC 1920
 QY 1921 AGACNTCTGAGAGCCCTGCGCNGAGTTCTCAACTTAAGGAGGCGNNNCTTAATTG 1980
 DB 1921 AGACNTCTGAGAGCCCTGCGCNGAGTTCTCAACTTAAGGAGGCGNNNCTTAATTG 1980
 QY 1981 AGAGAAA 1987
 DB 1981 AGAGAAA 1987

RESULT 6
 US-10-104-047-972
 ; Sequence 972, Application US/10104047
 ; Publication No. US20030236392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTION
 ; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
 ; FILE REFERENCE: H1-A0105

QY 33 GCCCATCCCAACCAAGCCCGGTGGGCAAGCAATCAACCTTAGAGAGCCCTATGA 92
 DB 539 GGCCTTCCCAAGCCCGCCCGGTGGGCAAGCAATCAACCTTAGAGAGCCCTATGA 598
 QY 93 GTTGGCGGTGAGCTGAGACCTTCTCACTGAAAGCATATATGTACCAACCTCCAAACA 152
 DB 599 GTTGGCGGTGAGCTGAGACCTTCTCACTGAAAGCATATATGTACCAACCTCCAAACA 152
 QY 153 AACCAATGAGGTGGGCGGTGAGAAAGCTGGCGCTGAAAGGACACTGTACATGAAA 212
 DB 656 CAACCAATGAGGTGGGCGGTGAGAAAGCTGGCGCTGAAAGGACACTGTACATGAAA 212
 QY 213 CACCTTACGACTACACAGTGTCCAGATCTGCGGAGACGTGAGACCCCGAGCGGTGA 272
 DB 714 TTGCT 762
 QY 273 CCTGGGCTACTGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 332
 DB 763 CTTGGGCT 819
 QY 333 CAGAACAGTGCAGACAGACCTTCCGAGCGGAGATCAAAATCTAGTGTCTCTCTCTCT 352
 DB 820 CAGAACAGTGCAGACAGACCTTCCGAGCGGAGATCAAAATCTAGTGTCTCTCTCTCT 879
 QY 393 TTTCCCTGCTCCCGCCCGCCAGCGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 452
 DB 880 TTTCCCTGCTCCCGCCCGCCAGCGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 939
 QY 453 TCCAGACATCTCAGCCAGGCTCTAGCCCTTCAAGCCCGCAGCAGCCAGGAGGAGGAGG 512
 DB 940 TCCAGACATCTCAGCCAGGCTCTAGCCCTTCAAGCCCGCAGCAGCCAGGAGGAGGAGG 999
 QY 513 CTCCCAACTAGGAGGCTCTCAGCTCTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 572
 DB 1000 CTCCCAACTAGGAGGCTCTCAGCTCTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1059
 QY 573 ATGATGCCAGATTTTCAAGTTTGGGCTCTCAGCTCTTATTCAGAGTACAGGAGGAGG 632
 DB 1060 ATGATGCCAGATTTTCAAGTTTGGGCTCTCAGCTCTTATTCAGAGTACAGGAGGAGG 1119
 QY 633 CAGGGAAGGAGATCTTAAGAACCCACTGTGGGTCAAGGGGATGGGACAGAGGACATTA 692
 DB 1120 CAGGGAAGGAGATCTTAAGAACCCACTGTGGGTCAAGGGGATGGGACAGAGGACATTA 1179
 QY 693 TGGGCAAGCTCTGAGAGACAGACAGACAGACCAACCTCTGATCTTAAGGATCTCTGAG 752
 DB 1180 TGGGCAAGCTCTGAGAGACAGACAGACAGACCAACCTCTGATCTTAAGGATCTCTGAG 1239
 QY 753 GGCAGAGGAGACAGAGGACCTGGAACCTCTTGGCCAAAGGAGGTGGAGAGACAGAGGGA 812
 DB 1240 GGCAGAGGAGACAGAGGACCTGGAACCTCTTGGCCAAAGGAGGTGGAGAGACAGAGGGA 1299
 QY 813 AGGTCAAGGCAAGGAGGCTTATCTAAGTGAAGTAAATTTGCGGAGGAGGCTCAGCAAGGC 872
 DB 1300 AGGTCAAGGCAAGGAGGCTTATCTAAGTGAAGTAAATTTGCGGAGGAGGCTCAGCAAGGC 1359
 QY 873 AAGAGAGACAGGCGTGAAGGTAACCTTCCCTCTTCAAGAGCTTCAAGGCCCAAGCCAGC 932

Query Match 83.6%; Score 1660.6; DB 15; Length 2410;
 Best Local Similarity: 95.6%; Pred. No. 0;
 Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;


```

RESULT 11
US-09-873-319-23/C
; Sequence 23, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 AA045503
US-09-873-319-23

Query Match      20.2%; Score 402; DB 10; Length 402;
Best Local Similarity 100.0%; Pred. No. 5.6e-119; Indels 0; Gaps 0;
Matches 402; Conservative 0; Mismatches 0;

QY 1521 TGGGCACTGAGAGATTTCCCTCATTTCTACTCTCTAGAGCTATGACCCCTCCAC 1580
DB 402 TGGGCACTGAGAGATTTCCCTCATTTCTACTCTCTAGAGCTATGACCCCTCCAC 343
QY 1581 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1640
DB 342 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 283
QY 1641 GGGACCTTCATCAGCTTACCTTGGGAAATGCTGCTGCCCCAGTACTTGGTTT 1700
DB 282 GGGACCTTCATCAGCTTACCTTGGGAAATGCTGCTGCCCCAGTACTTGGTTT 223
QY 1701 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1760
DB 222 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 163
QY 1761 CCAGGGCAAGTCAAGCTCTCCCTCCATGCTCTCTGCTGAGTGTGCTTGGGCT 1820
DB 162 CCAGGGCAAGTCAAGCTCTCCCTCCATGCTCTCTGCTGAGTGTGCTTGGGCT 103
QY 1821 CTCACTCCCACTCTGAGGCTTGGGGGAGGACTGGGGAGGGGGCGGTGGAGAGCC 1880
DB 102 CTCACTCCCACTCTGAGGCTTGGGGGAGGACTGGGGAGGGGGCGGTGGAGAGCC 43
QY 1881 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGCTTACAG 1922
DB 42 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGCTTACAG 1

RESULT 12
US-09-873-367C-532/C
; Sequence 532, Application US/09873367C
; Publication No. US20030168839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets

```

```

; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 532
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-532

Query Match      20.2%; Score 402; DB 10; Length 402;
Best Local Similarity 100.0%; Pred. No. 5.6e-119; Indels 0; Gaps 0;
Matches 402; Conservative 0; Mismatches 0;

QY 1521 TGGGCACTGAGAGATTTCCCTCATTTCTACTCTCTAGAGCTATGACCCCTCCAC 1580
DB 402 TGGGCACTGAGAGATTTCCCTCATTTCTACTCTCTAGAGCTATGACCCCTCCAC 343
QY 1581 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1640
DB 342 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 283
QY 1641 GGGACCTTCATCAGCTTACCTTGGGAAATGCTGCTGCCCCAGTACTTGGTTT 1700
DB 282 GGGACCTTCATCAGCTTACCTTGGGAAATGCTGCTGCCCCAGTACTTGGTTT 223
QY 1701 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1760
DB 222 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 163
QY 1761 CCAGGGCAAGTCAAGCTCTCCCTCCATGCTCTCTGCTGAGTGTGCTTGGGCT 1820
DB 162 CCAGGGCAAGTCAAGCTCTCCCTCCATGCTCTCTGCTGAGTGTGCTTGGGCT 103
QY 1821 CTCACTCCCACTCTGAGGCTTGGGGGAGGACTGGGGAGGGGGCGGTGGAGAGCC 1880
DB 102 CTCACTCCCACTCTGAGGCTTGGGGGAGGACTGGGGAGGGGGCGGTGGAGAGCC 43
QY 1881 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGCTTACAG 1922
DB 42 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGCTTACAG 1

RESULT 13
US-09-918-995-3808
; Sequence 3808, Application US/0918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3808
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

LOCATION: (1)...(486)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-3808

Query Match 20.1%; Score 399.4; DB 10; Length 486;
Best Local Similarity 99.8%; Pred. No. 4,2e-118;
Matches 400; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 640 GGAAGATCTAAAGAACCCATGTGGGTCAAGGGGAAATGGGACGACGACATATGGGCA 699
DB 85 GGAAGATCTAAAGAACCCATGTGGGTCAAGGGGAAATGGGACGACGACATATGGGCA 144
QY 700 GCTCTGAGAGACAGACAGACAGACAAACCTCTGATCTATGAAATCTTTCAGGGCAAG 759
DB 145 GCTCTGAGAGACAGACAGACAGACAAACCTCTGATCTATGAAATCTTTCAGGGCAAG 204
QY 760 GACACAGGAGACCTGGAAACCTCTTGGCCAAAGGAGTGGAGAGACAGAGGAAAGTCA 819
DB 205 GACACAGGAGACCTGGAAACCTCTTGGCCAAAGGAGTGGAGAGAGAGAGAGTCA 264
QY 820 AGGCAAGGGTGCCTATCTATGAGAGACTAATGGCCGAGGGCTCAGACAGGCCAGAGGA 879
DB 265 AGGCAAGGGTGCCTATCTATGAGAGACTAATGGCCGAGGGCTCAGACAGGCCAGAGGA 324
QY 880 GACAGCCGTGACGGTAAACTTCCCTCTACCAAGCTCCAAAGCCCAAGCCAGAGCAG 939
DB 325 GACAGCCGTGACGGTAAACTTCCCTCTACCAAGCTCCAAAGCCCAAGCCAGAGCAG 384
QY 940 CTGCTGCGCCACCCCGCCCGCCAGCAGCTGTGTGCAAGGCGACCAATGGCAAT 959
DB 385 CTGCTGCGCCACCCCGCCCGCCAGCAGCTGTGTGCAAGGCGACCAATGGCAAT 444
QY 1000 CTGTATATAGATGGGGTTTCCCATACAGCTGGTCTGA 1040
DB 445 CTGTATATAGATGGGGTTTCCCATACAGCTGGTCTGA 485

RESULT 14
US-09-918-995-4147
Sequence 4147, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4147
LENGTH: 400
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-4147

Query Match 18.9%; Score 375; DB 10; Length 400;
Best Local Similarity 99.2%; Pred. No. 3e-110;
Matches 397; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1088 GCAAGCCCAAGTGGGGTGGGGTGGGGCTGGCTCCCTCCCAAGAGGCTGTGTC 1147
DB 1 GCAAGCCCAAGTGGGGTGGGGTGGGGCTGGCTCCCTCCCAAGAGGCTGTGTC 60
QY 1148 NTGGGGCTGCTCCATGACAGACAGATCACTTAACAGAGATGAAGCCAGGGCATGATG 1207
DB 61 TTGGGGCTGCTCCATGACAGACAGATCACTTAACAGAGATGAAGCCAGGGCATGATG 120
QY 1208 GGGCTTTGGGTCTCAGAGTTGAAGCCCGCTTTTGGCACTTCCCTCC-GGAGTCA 1266
DB 121 GGGCTTTGGGTCTCAGAGTTGAAGCCCGCTTTTGGCACTTCCCTCCCGGGCATCA 180

QY 1267 GCTCTCATCATCCCCCTCTTATATATGATCTATAGGCTGGTGTGTACAC-AC 1325
DB 181 GCTCTCATCATCCCCCTCTTATATATGATCTATAGGCTGGTGTGTACACAC 240
QY 1326 ACACCCCATGCTGTCTCTCAATACATGATCATGATGATGAGGCAATTCAGA 1385
DB 241 ACACCCCATGCTGTCTCTCAATACATGATCATGATGATGAGGCAATTCAGA 300
QY 1386 GCTTCTCAATCAGATTTACATCTCATTTTCATTACGGGGAACATCCCGACCA 1445
DB 301 GCTTCTCAATCAGATTTACATCTCATTTTCATTACGGGGAACATCCCGACCA 360
QY 1446 CTGAGTGTGTGCTTGTCACTGAGGTTAGTCAACC 1485
DB 361 CTGAGTGTGTGCTTGTCACTGAGGTTAGTCAACC 400

RESULT 15
US-09-864-761-2033/c
Sequence 2033, Application US/03864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2033
LENGTH: 458
TYPE: DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 11:30:51 ; Search time 77 Seconds

(without alignments)
708.141 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPTLYIDGVFPQLVREKLH.....LSQIRPTISIFINGETSPSH 152

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: AGeneseq_23sep04:*
2: geneseqp1980s:*
3: geneseqp1980s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003s:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	846	99.8	152	6	ABO23236 Human bre
2	846	99.8	152	6	ADP66735 Novel hum
3	846	99.8	152	8	ADO60245 Human NOV
4	836.5	98.6	151	4	AAB61132 Human NOV
5	836.5	98.6	151	7	ADM56371 Human cel
6	739	87.1	136	8	AD119772 Human NOV
7	500.5	59.0	102	2	AAV48502 Human bre
8	312	36.8	101	6	AAB61152 Protein e
9	312	36.8	101	6	ABO23256 Human bre
10	312	36.8	101	8	ADP66803 Novel hum
11	312	36.8	101	8	ADP66803 Novel hum
12	305	36.0	101	7	ADM56440 Human cel
13	296	34.9	101	8	AD119840 Human NOV
14	113.5	13.4	703	5	ABR05728 Human tita
15	95	11.2	247	7	ADB64988 Human pro
16	90.5	10.7	502	8	ABO59296 Human gen
17	90.5	10.7	665	8	ADN99681 Novel hum
18	88	10.4	200	5	ADK34802 Novel hum
19	88	10.4	1745	4	AAQ77793 Human pro
20	88	10.4	1745	5	ABR97234 Novel hum
21	88	10.4	1745	8	ADQ19841 Human sof
22	85	10.0	181	8	ADG22767 Cyanophag
23	84.5	10.0	232	5	ABP41247 Human ova
24	84.5	10.0	334	7	ABO67883 Pseudomon
25	83.5	9.8	631	5	ABG32473 Human pro

26	83	9.8	147	4	AAU32381 Novel hum
27	82.5	9.7	68	4	AAQ74508 Human col
28	82.5	9.7	884	2	AAW59666 Amino aci
29	82.5	9.7	884	2	AAW94074 Human G-P
30	81.5	9.6	457	7	ABO80925 Pseudomon
31	81	9.6	1739	4	AAQ77792 Murine pr
32	80.5	9.5	1554	4	ABR11890 Human pro
33	80.5	9.5	3298	4	AAE03657 Human ext
34	80.5	9.5	3312	7	ADJ69693 Human hea
35	80.5	9.5	3312	8	ADO29247 Human GPC
36	80.5	9.5	3312	8	ADO17551 Human sof
37	80.5	9.5	3317	8	ADH18895 Human cel
38	80.5	9.5	4115	5	ABP53570 Human NOV
39	80	9.4	361	7	ADP94865 Human gen
40	80	9.4	365	8	ADL71718 Novel hum
41	80	9.4	542	7	ADP94893 Human gen
42	80	9.4	542	8	ADL71719 Novel hum
43	80	9.4	583	7	ADJ70094 Human hea
44	80	9.4	920	6	ABP70827 Human C1Q
45	80	9.4	1016	3	ABR41524 Human ORF

ALIGNMENTS

RESULT 1
ID ABO23236 standard, protein, 152 AA.
XX
AC ABO23236;
XX
DT 03-SEP-2003 (first entry)
XX
DE Human breast tumour associated protein 47-like polypeptide NOV4.
XX
XX Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
XX pancreatic cancer; uterine cancer; organ transplantation disorder;
XX cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
XX ischaemic heart disease; haemorrhage; peripheral vascular disease;
XX thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;
XX tissue regeneration; wound healing; hyperproliferative disorder;
XX psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
XX Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
XX ALS; peripheral neuropathy; nervous system tumour; neurocystic; tremor;
XX neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;
XX gene therapy; epilepsy; breast tumour associated protein 47.
XX
XX Homo sapiens.
OS
XX
XX US2003027158-A1.
PN
XX
PD 06-FEB-2003.
XX
XX 15-OCT-2001; 2001US-00977418.
XX
XX 03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkete RA, Fernandes E, Herrman J, Vernet C;
PI WPI: 2003-492028/46.
DR N-PSDB; ACDB0257.
XX
XX New nucleic acid sequence encoding a human breast tumor-associated
PT protein 47-like polypeptide, useful for treating cardiovascular
PT disorders, neural disorders, diabetes mellitus and cancers.
XX

PS Claim 1; Page 16-17; 100pp; English.

XX The invention relates to a new isolated NOV4 nucleic acid. The nucleic acid is useful for identifying a compound that binds the nucleic acid. The nucleic acid is useful in gene therapy, in screening assays, in detection assays e.g. chromosomal mapping, cell and tissue typing and forensic biology, predictive medicine e.g. diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenomics and methods of treatment including therapeutic and prophylactic. The nucleic acid is also useful for expressing NOVX protein. The nucleic acid is also useful to provide polynucleotide reagents e.g. labelled probes that are useful in an in situ hybridisation technique, for identifying a specific tissue (for example brain tissue) and for use in forensic science. The nucleic acid is also useful for mapping genes on a chromosome and thus locating gene regions associated with genetic disease, identifying an individual from a minute biological sample and to aid in forensic identification of biological sample. The nucleic acid is also useful for treating cancer, especially cancers of the breast, colon, lung, pancreas or uterus, or a melanoma or sarcoma. The nucleic acid is also useful for treating disorders related to organ transplantation, cardiovascular diseases, atherosclerosis, ischaemic heart disease, haemorrhage, diabetes mellitus, peripheral vascular disease, thrombosis, hypertension and systemic lupus erythematosus. NOVX protein encoded by the nucleic acid is useful for regulating haematopoiesis, for regeneration of bone, cartilage, tendon ligament and/or nerve tissue growth or regeneration and for wound healing. The nucleic acid is also useful for treating infections, hyperproliferative disorders e.g. psoriasis, and neural disorders including Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumours of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury and other neuropathies, epilepsy and/or tremors. The present sequence represents the amino acid sequence of a human breast tumour associated protein 47-like polypeptide

XX Sequence 152 AA;

Query Match 99.8%; Score 846; DB 6; Length 152;

Best Local Similarity 100.0%; Pred. No. 3,3e-84; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELYIDGVFPPIOLVREKHEPVALRLIGPPGKAKMGWGMGMSFLPQACVXGAAPMOT 60
DB 1 MPELYIDGVFPPIOLVREKHEPVALRLIGPPGKAKMGWGMGMSFLPQACVXGAAPMOT 60
QY 61 GSNRDSGQMDALPPRGWTPASCHLPURQSLHPSPLIYESIGVCVTHPRCPNS 120
DB 61 GSNRDSGQMDALPPRGWTPASCHLPURQSLHPSPLIYESIGVCVTHPRCPNS 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 2

AD66735 ADF66735 standard; protein; 152 AA.

AC ADF66735;

DT 12-FEB-2004 (first entry)

XX Novel human protein NOV4.

XX cytosolic; hepatotropic; vulnery; antipsoriatic; osteopathic;
XX antiarthritic; antiatherosclerotic; haemostatic; vasotropic;
XX thrombolytic; antidiabetic; hypotensive; dermatological;
XX immunosuppressive; antinflammatory; immunostimulant; fungicide;
XX virucide; protozoacide; neuroprotective; antineumatic; antiarthritic;
XX antistatic; antiparkinsonian; nootropic; anticonvulsant;
XX NOVX modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
XX psoriasis; tissue hypertrophy; osteoarthritis;
XX atherosclerotic plaque formation; haemorrhage; ischaemic disease;
XX thrombosis; diabetes mellitus; hypertension; hypothyroidism;

KW immune deficiency; severe combined immunodeficiency; SCID; infection;
KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
KW multiple sclerosis; systemic lupus, erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KW autoimmune thyroiditis; insulin dependent diabetes mellitus;
KW myasthenia gravis; graft- versus-host disease;
KW autoimmune inflammatory eye disease; asthma; haematopoiesis;
KW tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
KW periodontal disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KW human.

OS Homo sapiens.

PN US2003199103-A1.

PD 23-OCT-2003.

PF 15-OCT-2001; 2001US-00977639.

PR 03-JUN-1999; 99US-0137322P.

PR 16-MAR-2000; 2000US-0189810P.

PR 22-MAR-2000; 2000US-0191158P.

PR 30-MAR-2000; 2000US-0193086P.

PR 03-MAY-2000; 2000US-0201388P.

PR 31-MAY-2000; 2000US-00584411.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Fernandes E, Herrman J, Vernet C;

PT WPI; 2004-021196/02.

PR N-PSDB; ADF66734.

XX Novel substantially NOVX polypeptide useful for diagnosing, preventing

PT and treating diseases e.g., cancer, multiple sclerosis, systemic lupus

PR erythematosus.

PT Disclosure; SEQ ID NO 8; 165pp; English.

XX The invention describes a substantially purified polypeptide (I) having
XX amino acid sequence chosen from a fully defined NOVX sequence (SI) of 708
XX amino acids as given in the specification, or polypeptide having one or
XX more conservative amino acid substitutions of (SI) or mutant or variant
XX of (SI). (I) having (SI) is useful for diagnosing a pathological
XX condition associated with (I) or its activity in a subject e.g. cancer.
XX (I) useful in treatment of cancer, hyperproliferative diseases,
XX cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis,
XX atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal
XX disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I)
XX is useful in treatment of various immune deficiencies and disorders such
XX as severe combined immunodeficiency (SCID), bacterial infection, viral
XX infection such as herpes viral infection, protozoan infection such as
XX malaria, fungal infection such as candidiasis. (I) is also useful in
XX treating autoimmune disorders such as connective tissue disease, multiple
XX sclerosis, systemic lupus, erythematosus, rheumatoid arthritis,
XX autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune
XX thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis,
XX graft- versus-host disease and autoimmune inflammatory eye disease and
XX asthma. (I) useful in regulation of haematopoiesis, regeneration and
XX tissue growth of bone, cartilage, tendon, ligament and useful for wound
XX healing and tissue repair. (I) is also useful in treatment of burns,
XX incisions and ulcers. (I) also useful in treatment of periodontal
XX disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective
XX antitumour and antiinflammatory activity. This is the amino acid sequence
XX of a novel human NOVX protein.

XX Sequence 152 AA;

Query Match 99.8%; Score 846; DB 8; Length 152;

Best Local Similarity 100.0%; Pred. No. 3,3e-84; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELYIDGVFPPIQLVREKLEHETPAVRLILGPPGKAKWGMGWSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVFPPIQLVREKLEHETPAVRLILGPPGKAKWGMGWSFSLPQACVXGAAPMOT 60
QY 61 GSPNPDGSGMDGALGPRGWTTPASCHLPLRQSAALHPSPSLIYESIGSVCTVTHPRCPSPN 120
DB 61 GSPNPDGSGMDGALGPRGWTTPASCHLPLRQSAALHPSPSLIYESIGSVCTVTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152

RESULT 3
ADO60245 ID ADO60245 standard; protein; 152 AA.
XX
AC ADO60245;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human NOVA4 protein.
XX
KM Human; NOVA4 protein; cancer; hyperproliferative disease; cirrhosis;
KM keloid; psoriasis; tissue hypertrophy; osteoarthritis;
KM atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
KM renal disease; thrombosis; diabetes mellitus; hyperextension;
KM hypothyroidism; severe combined immunodeficiency; SCID; infection;
KM malaria; candidiasis; autoimmune disorder; connective tissue disease;
KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KM autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KM autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
KM autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
KM periodontal disease; Alzheimer's disease; Parkinson's disease;
KM Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KM haematopoiesis; wound healing; tissue repair; antitumor;
KM antinflammatory.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= Signal_peptide
FT Protein 26..152
FT /note= "Mature NOVA4 protein"
FT Misc-difference 53
FT /note= "Encoded by CNT"
XX
FT US2003134430-A1.
XX
PD 17-JUL-2003.
XX
PF 15-OCT-2001; 2001US-0097751.
XX
PR "03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkels RA, Fernandes E, Herrman J, Vernet C;
XX
DR WPI; 2004-068928/07.
XX
DR N-PSDB; ADO60244.
XX
PT Novel substantially purified NOVA4 polypeptide for treating severe
PT combined immunodeficiency, candidiasis, cancer, asthma, multiple
XX sclerosis, systemic lupus erythematosus.
XX

PS Disclosure; SEQ ID NO 8; 155pp; English.
XX
CC The invention relates to human NOVA4 polypeptides and polynucleotides.
CC NOVA4 sequences are useful in the treatment of cancer, hyperproliferative
CC diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
CC osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
CC heart or renal disease, thrombosis, diabetes mellitus, hyperextension,
CC hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
CC Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
CC and disorders such as severe combined immunodeficiency (SCID), bacterial
CC infection, viral infection e.g. herpes viral infection, protozoan
CC infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
CC disorders such as connective tissue disease, multiple sclerosis, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune thyroiditis, insulin
CC inflammation, Guillain-Barre syndrome, autoimmune hypothyroiditis, insulin
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease. The invention is useful in
CC regeneration and tissue growth of bone, cartilage, tendon, ligament,
CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
CC the invention also exhibit antitumor and antinflammatory activities. The
CC present sequence is human NOVA4 protein.
XX
SQ Sequence 152 AA;
XX
Query Match 99.8%; Score 846; DB 8; Length 152;
Best Local Similarity 100.0%; Pred. No. 3.3e-84;
Matches 152; Conservative 100; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPELYIDGVFPPIQLVREKLEHETPAVRLILGPPGKAKWGMGWSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVFPPIQLVREKLEHETPAVRLILGPPGKAKWGMGWSFSLPQACVXGAAPMOT 60
QY 61 GSPNPDGSGMDGALGPRGWTTPASCHLPLRQSAALHPSPSLIYESIGSVCTVTHPRCPSPN 120
DB 61 GSPNPDGSGMDGALGPRGWTTPASCHLPLRQSAALHPSPSLIYESIGSVCTVTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152

RESULT 4
AAB61132 ID AAB61132 standard; protein; 151 AA.
XX
AC AAB61132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human NOVA4 protein.
XX
KM Human; NOVA4; antinflammatory; cyostatic; neuroprotective;
KM cerebroprotective; immunomodulator; vulnerrary; vasculoprotic; gene therapy;
KM hyperplasia; tumour; reestenosis; psoriasis; Dupuytren's contracture;
KM diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN WO200075321-A2.
XX
PD 14-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US015303.
XX
PR 03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 31-MAY-2000; 2000US-00137322.
XX
PA (CURA-) CURAGEN CORP.
XX

PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI: 2001-102403/11.
 DR N-PSDB; AAF27852.
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing.
 XX
 PS Claim 1; Page 26-27; 194pp; English.
 CC The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasia, tumors,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)
 CC
 XX Sequence 151 AA;
 SQ
 Query Match 98.6%; Score 836.5; DB 4; Length 151;
 Best Local Similarity 99.3%; Pred. No. 3,6e-83;
 Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 MPELYIDGVFPPIQLVREKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACVXGAAPMOT 60
 Db 1 MPELYIDGVFPPIQLVREKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACV-GAAPMOT 59
 QY 61 GSPNRDGSQMDGALGPRGWTASCHLPIRQSLHPSLSLYESIGSVCTTHPRCPSN 120
 Db 60 GSPNRDGSQMDGALGPRGWTASCHLPIRQSLHPSLSLYESIGSVCTTHPRCPSN 119
 QY 121 TCHYHWRNPSELSQIRFTISIFINGETSPSH 152
 Db 120 TCHYHWRNPSELSQIRFTISIFINGETSPSH 151
 RESULT 5
 ADMS6371
 ID ADM56371 standard; protein; 151 AA.
 XX
 AC ADM56371;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human cell adhesion molecule NOV4 #1.
 XX
 KW Human: cell adhesion molecule; NOVX; Cancer; leukaemia; lymphoma;
 KW melanoma; neurological disorder; epilepsy; Alzheimer's disease;
 KW ischaemic cerebrovascular disease; stroke; Alzheimer's disease;
 KW Pick's disease; vesicular transport disorder; cystic fibrosis;
 KW diabetes mellitus; Grave's disease; goiter; gastrointestinal disorder;
 KW ulcerative colitis; gastric ulcer; duodenal disorder; autoimmune disease;
 KW allergic reaction; autoimmune haemolytic anaemia; rheumatoid arthritis;
 KW viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infections.
 XX
 OS Homo sapiens.
 XX
 PN US2003082554-A1.
 XX
 PD 01-MAY-2003.

XX
 PF 15-OCT-2001; 2001US-00977033.
 XX
 PR 03-JUN-1999; 99US-0137322P.
 XX
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191159P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CUPRA-) CUPRAGEN CORP.
 XX
 PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI: 2003-616079/58.
 DR N-PSDB; ADM56372.
 PT New nucleic acids encoding human cell adhesion molecule-like proteins,
 PT useful for treating e.g. cancers, neurological disorders, viral,
 PT bacterial, fungal, helminthic and protozoal infections.
 XX
 PS Example 3; SEQ ID NO 8; 78pp; English.
 CC The invention relates to an isolated nucleic acid encoding a human cell
 CC adhesion molecule-like protein, comprising a sequence encoding a
 CC polypeptide having a sequence appearing as ADM56387, a sequence at least
 CC 90% identical to the nucleic acid, a sequence encoding a polypeptide
 CC having conservative amino acid substitutions to the protein or a fragment
 CC comprising at least 20 nucleotides. Also included are an oligonucleotide
 CC sequence that is complementary to (and hybridises under stringent
 CC conditions with) the nucleic acid (or a portion of it), a vector
 CC comprising the nucleic acid, a cell comprising the vector, a
 CC pharmaceutical composition comprising the nucleic acid and a
 CC process for identifying a compound that binds the nucleic acid, and a
 CC compound identified by the process. Disclosed as new are the cDNA and
 CC proteins for novel cell adhesion molecules (termed NOVX, being NOV1-23).
 CC The NOVX polypeptide, nucleic acid or antibody are useful in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy,
 CC ischaemic cerebrovascular disease, stroke, Alzheimer's disease or Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or goiter), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The polypeptides can be used as immunogens to
 CC produce antibodies and as vaccines. The sequences may further be used in
 CC chromosome mapping, identifying individual from minute biological samples
 CC (tissue typing), and in forensic identification of a biological sample.
 CC NOTE: The authors have mis-labelled the sequences as they appear on pages
 CC 12-28 of the patent, it is clear from table 3, the examples and the
 CC claims that the SEQ ID numbers for the cDNAs should be the odd numbers
 CC from 1-45 and the proteins should be the even numbers from 2-46. The
 CC present sequence represents a NOVX cell adhesion molecule of the
 CC invention.
 XX
 SQ Sequence 151 AA;
 QY Query Match 98.6%; Score 836.5; DB 7; Length 151;
 Best Local Similarity 99.3%; Pred. No. 3,6e-83;
 Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 Db 1 MPELYIDGVFPPIQLVREKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACV-GAAPMOT 59
 QY 61 GSPNRDGSQMDGALGPRGWTASCHLPIRQSLHPSLSLYESIGSVCTTHPRCPSN 120
 Db 60 GSPNRDGSQMDGALGPRGWTASCHLPIRQSLHPSLSLYESIGSVCTTHPRCPSN 119

QY 121 TOHHMLRPNSLSQIRFTISIFINGETSPSH 152
 DB 120 TOHHMLRPNSLSQIRFTISIFINGETSPSH 151

RESULT 6

AD119772
 ID AD119772 standard; protein; 136 AA.

AC AD119772;
 DT 22-APR-2004 (first entry)
 XX

DE Human NOV4 protein.

XX Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 XX infection; anorexia; cancer; cardiovascular disease; hypertension;
 XX atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
 XX Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
 XX haematopoietic disorder; inflammatory skin disorder; asthma;
 XX dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 XX haematopoiesis; wound healing; angiogenesis; chromosome mapping;
 XX tissue typing; preventive medicine; pharmacogenomic; gene therapy;
 XX anorectic; cardiac; viricide; antibacterial; fungicide; protozoacide;
 XX neurotropic; neuroprotective; dermatologically; human.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..54
 FT MISC-difference 52..53

FT Protein /note= "Encoded by GTTCNTGGG"
 FT /note= 55..136

FT MISC-difference 60..61 /note= "Human mature NOV protein"
 FT /note= "Encoded by bases from 1171 to 1221"

PN US2004002134-A1.

PD 01-JAN-2004.

PF 15-OCT-2001; 2001US-00977819.

PR 03-JUN-1999; 99US-0137322P.

PR 16-MAR-2000; 2000US-0189810P.

PR 22-MAR-2000; 2000US-0191158P.

PR 30-MAR-2000; 2000US-0193086P.

PR 03-MAY-2000; 2000US-0201388P.

PR 31-MAY-2000; 2000US-00584411.

XX (CUPA-) CURAGEN CORP.

PI Shinkets RA, Fernandes ER, Herrman JL, Vernet CAM;

DR WPI; 2004-070737/07.

DR N-PSTB; AD119771.

PT New NOVX nucleic acids encoding human K1A0768 protein-like and human

PT protein PRO-228 polypeptides, useful for treating NOVX-associated

PT disorders.

XX

XX

XX

CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV protein.
 XX
 XX

SO Sequence 136 AA;

Query Match 87.1%; Score 739; DB 8; Length 136;
 Best Local Similarity 89.5%; Pred. No. 1.5e-72;
 Matches 136; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MPELYIDGVFPQLVREKLEHPVALRLGPPGKAKMGWGWMSFSLPQACVXGAAPNOT 60

DB 1 MPELYIDGVFPQLVREKLEHPVALRLGPPGKAKMGWGWMSFSLPQACV-GAAPNOT 59

QY 61 GSPNRDGSQMDGALGPRGWPASCHLPKQSAHPSPSLYESTGVCVTTTHPRCSN 120

DB 60 -----GPRGWPASCHLPKQSAHPSPSLYESTGVCVTTTHPRCSN 104

QY 121 TOHHMLRPNSLSQIRFTISIFINGETSPSH 152

DB 105 TOHHMLRPNSLSQIRFTISIFINGETSPSH 136

RESULT 7

AA48502
 ID AA48502 standard; protein; 102 AA.

AC AA48502;

DT 08-DEC-1999 (first entry)

DE Human breast tumour-associated protein 47.

XX Expressed sequence tag; EST; human; breast; cancer; cytostatic;

XX medicaments; gene therapy; treatment; fat metabolism.

OS Homo sapiens.

PN DE19813835-A1.

PD 23-SEP-1999.

PF 20-MAR-1998; 98DE-01013835.

PR 20-MAR-1998; 98DE-01013835.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenchal A;

DR WPI; 1999-528979/45.

DR N-PSTB; AA233583.

PT Human nucleic acid sequences and protein products from normal breast

PT tissue, useful for breast cancer therapy.

XX Claim 28; 179; 206BP; German.

CC This invention describes novel human nucleic acid sequences from normal

CC breast tissue which have cytoskeletal activity. The nucleic acid sequences

CC can be used to produce and isolate full-length gene sequences. They can

CC be used to express proteins, which can be used as tools to find a

CC activity against breast cancer. The sequences can be used in sense or

CC antisense form. They are especially useful for medicines associated with

CC fat metabolism. AA48456-Y48539 represent protein fragments encoded by

CC the expressed sequence tags described in the method of the invention

XX Sequence 102 AA;

Query Match 59.0%; Score 500.5; DB 2; Length 102;
 Best Local Similarity 90.0%; Pred. No. 1,3e-46;
 Matches 90; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 1 MPHLYIDGVPIQLVREKHEHPAVLRLLGPPGKATKMGWGSFSLPQACVYGAAPMOT 60
 DB 1 MPHLYIDGVPIQLVREKHEHPAVLRLLGPPGKATKMGWGSFSLPQACVYGAAPMOT 60

QY 61 GSPNRDSQGMDSGALGPRGWTTPASCHLPLRQSLHSPSLI 100
 DB 61 GSPNRDSQGMDSGALGPRGWTTPASCHLPLRAVS---SPSL 97

RESULT 8
 AAB61152
 ID AAB61152 standard; protein; 101 AA.
 AC AAB61152;
 DT 30-MAR-2001 (first entry)
 DE Protein encoded by human NOVA cDNA clone 3189601.
 KW Human; NOVA; antiinflammatory; cytosolic; neuroprotective;
 KW cerebroprotective; immunomodulator; vasotrophic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 OS Homo sapiens.
 PN MO200075321-A2.
 PD 14-DEC-2000.
 PF 01-JUN-2000; 2000MO-US015303.
 XX
 PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 31-MAY-2000; 2000US-00137322.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 DR WPI; 2001-102403/11.
 XX
 PT New NOVA polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing.
 XX
 PS Example 3; Page 174; 194pp; English.
 XX
 CC The present sequence is a new isolated polypeptide (NOVA). The NOVA
 CC polypeptides, NOVA nucleic acids, and anti-NOVA antibodies are useful for
 CC treating or preventing NOVA-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVA polypeptide or nucleic acid.
 CC These NOVA-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVA polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVA is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)
 CC
 XX Sequence 101 AA;

Query Match 36.8%; Score 312; DB 4; Length 101;
 Best Local Similarity 60.0%; Pred. No. 5.4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPHLYIDGVPIQLVREKHEHPAVLRLLGPPGKATKMGWGSFSLPQACVYGAAPMOT 60
 DB 1 MPHLYIDGVPIQLVREKHEHPAVLRLLGPPGKATKMGWGSFSLPQACVYGAAPMOT 60

QY 61 GSPNRDSQGMDSG---ALGPRGWTTPASCHLPLRQSLHSPSLI 101
 DB 61 DHLTMEFARAWMGLWLVGVQ---LLATFPQSQSLHSPSLI 101

RESULT 9
 ABO23256
 ID ABO23256 standard; protein; 101 AA.
 AC ABO23256;
 DT 03-SEP-2003 (first entry)
 DE Human breast tumour associated protein 47-like polypeptide NOVA product.
 KW Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
 KW pancreatic cancer; uterine cancer; organ transplantation disorder;
 KW cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
 KW ischaemic heart disease; haemorrhage; peripheral vascular disease;
 KW thrombosis; hypertension; systemic lupus erythematosus; haematoplasia;
 KW tissue regeneration; wound healing; hyperproliferative disorder;
 KW psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
 KW Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KW ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
 KW neuropathy; acute brain injury; peripheral nerve trauma; human; NOVA;
 KW gene therapy; epilepsy; breast tumour associated protein 47.
 XX
 OS Homo sapiens.
 PN US2003027158-A1.
 PD 06-FEB-2003.
 PF 15-OCT-2001; 2001US-00977418.
 XX
 PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 DR WPI; 2003-492028/46.
 XX
 PT N-PSDB; ACD40305.
 PT New nucleic acid sequence encoding a human breast tumor-associated
 PT protein 47-like polypeptide, useful for treating cardiovascular
 PT disorders, neural disorders, diabetes mellitus and cancers.
 XX
 PS Example 3; Page 93; 100pp; English.
 XX
 CC The invention relates to a new isolated NOVA nucleic acid. The nucleic
 CC acid is useful for identifying a compound that binds the nucleic acid.
 CC The nucleic acid is useful in gene therapy, in screening assays, in
 CC detection assays e.g. chromosomal mapping, cell and tissue typing and
 CC forensic biology, predictive medicine e.g. diagnostic assays, prognostic
 CC assays, monitoring clinical trials and pharmacogenomics and methods of
 CC treatment including therapeutic and prophylactic. The nucleic acid is
 CC also useful for expressing NOVA protein. The nucleic acid is also useful
 CC to provide polynucleotide reagents e.g. labelled probes that are useful

XX 15-JUL-2004 (first entry)
 XX Human NOV4 protein fragment.
 DE
 XX Human, NOVX protein; cancer; hyperproliferative disease; cirrhosis;
 KW keloid; psoriasis; tissue hypertrophy; osteoarthritis;
 KW atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
 KW renal disease; thrombosis; diabetes mellitus; hypertension;
 KW hypothyroidism; severe combined immunodeficiency; SCID; infection;
 KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
 KW autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
 KW autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
 KW pericardial disease; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
 KW haematopoiesis; wound healing; tissue repair; antitumor; anti-inflammatory;
 KW antiinflammatory.
 XX
 OS Homo sapiens.
 XX
 PN US2003134430-A1.
 PD
 XX 17-JUL-2003.
 PE
 XX 15-OCT-2001; 2001US-00977751.
 PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 DR WPI: 2004-068928/07.
 DR N-PSDB; ADO60312.
 PT Novel substantially purified NOVX polypeptide for treating severe
 PT combined immunodeficiency, candidiasis, cancer, asthma, multiple
 PT sclerosis, systemic lupus erythematosus.
 XX
 XX Example 3; SEQ ID NO 76; 155dp; English.
 PS
 CC The invention relates to human NOVX polypeptides and polynucleotides.
 CC NOVX sequences are useful in the treatment of cancer, hyperproliferative
 CC diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
 CC osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
 CC heart or renal disease, thrombosis, diabetes mellitus, hypertension,
 CC hypothyroidism, asthma, burns, incisions, ulcers, pericardial disease,
 CC Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
 CC and disorders such as severe combined immunodeficiency (SCID), bacterial
 CC infection, viral infection e.g. herpes viral infection, protozoan
 CC infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
 CC disorders such as connective tissue disease, multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
 CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
 CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
 CC and autoimmune inflammatory eye disease. The invention is useful in
 CC regeneration and tissue growth of bone, cartilage, tendon, ligament,
 CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
 CC the invention also exhibit antitumor and anti-inflammatory activities. The
 CC present sequence is human NOV protein fragment used in the
 CC exemplification of the invention.
 XX
 XX Sequence 101 AA;
 SQ

Query Match

36.8%; Score 312; DB 8; Length 101;

Best Local Similarity 60.0%; Pred. No. 5,4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;
 QY 1 MPHLYIOGVFPIQVREKHEETPAVRLILGPGKAKKMGWGSFSLPQACVYGAAPMOT 60
 DB 1 MPHLYIDVFPPIQVLRDHEETPAVRLILGPGKATWGWGWSFSLPVPVFLALPCQ 60
 QY 61 GSPNRDGSQMDG---ALGPRGWTPASCHLPLRQSLHPSPLI 101
 DB 61 DHELEMEARAMWGJMWLEVGPO---LLATPSSQSLHPSPLI 101
 RESULT 12
 ADM56440
 ID ADM56440 standard; protein; 101 AA.
 XX
 AC ADM56440;
 XX
 DT 03-JUN-2004 (first entry)
 DE
 XX Human cell adhesion molecule NOV4 #2.
 XX
 KW Human; cell adhesion molecule; NOVX; cancer; leukaemia; lymphoma;
 KW melanoma; neurological disorder; epilepsy;
 KW ischaemic cerebrovascular disease; stroke; Alzheimer's disease;
 KW Pick's disease; vesicular transport disorder; cystic fibrosis;
 KW diabetes mellitus; Grave's disease; goiter; gastrointestinal disorder;
 KW ulcerative colitis; gastric ulcer; duodenal disorder; autoimmune disease;
 KW allergic reaction; autoimmune haemolytic anaemia; rheumatoid arthritis;
 KW viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infections.
 XX
 OS Homo sapiens.
 XX
 PN US2003082554-A1.
 PD
 XX 01-MAY-2003.
 PE
 XX 15-OCT-2001; 2001US-00977703.
 PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 DR WPI: 2003-616079/58.
 DR N-PSDB; ADM56439.
 PT New nucleic acids encoding human cell adhesion molecule-like proteins,
 PT useful for treating e.g. cancers, neurological disorders, viral,
 PT bacterial, fungal, helminthic and protozoal infections.
 XX
 XX Example 3; SEQ ID NO 76; 78dp; English.
 PS
 CC The invention relates to an isolated nucleic acid encoding a human cell
 CC adhesion molecule-like protein, comprising a sequence encoding a
 CC polypeptide having a sequence appearing as ADM5387, a sequence at least
 CC 90% identical to the nucleic acid, a sequence encoding a polypeptide
 CC having conservative amino acid substitutions to the protein or a fragment
 CC comprising at least 20 nucleotides. Also included are an oligonucleotide
 CC sequence that is complementary to (and hybridises under stringent
 CC conditions with) the nucleic acid (or a portion of it), a vector
 CC comprising the nucleic acid, a cell comprising the vector, a
 CC pharmaceutical composition comprising the nucleic acid and a
 CC pharmaceutical carrier, a process for producing the polypeptide, a
 CC process for identifying a compound that binds the nucleic acid, and a
 CC compound identified by the process. Disclosed as new are the cDNA and

CC proteins for novel cell adhesion molecules (termed NOVX, being NOV1-23).
 CC The NOVX polypeptide, nucleic acid or antibody are useful in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or galter), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The polypeptides can be used as immunogens to
 CC produce antibodies and as vaccines. The sequences may further be used in
 CC chromosome mapping, identifying individual from minute biological samples
 CC (tissue typing), and in forensic identification of a biological sample.
 CC NOTE: The authors have mislabelled the sequences as they appear on pages
 CC 12-28 of the patent, it is clear from table 3, the examples and the
 CC claims that the SEQ ID numbers for the CDNA should be the odd numbers
 CC from 1-45 and the proteins should be the even numbers from 2-46. The
 CC present sequence represents a NOVX cell adhesion molecule of the
 CC invention.

CC Sequence 101 AA;

Query Match 36.0%; Score 305; DB 7; Length 101;
 Best Local Similarity 59.0%; Pred. No. 3.2e-25;
 Matches 62; Conservative 7; Mismatches 28; Indels 8; Gaps 2;

QY 1 MPEHYIGVFPFIQVREKIHETPAVRLIGPGRKAKMGWGMGMSFSLPQACVXGAAPMOT 60
 DB 1 MPEHYIGVFPFIQVREKIHETPAVRLIGPGRKAKMGWGMGMSFSLPQACVXGAAPMOT 60

QY 61 GSENRDGSQGMDS---ALGPRGWTTPASCHLPRLQSAALHPSPSLI 101
 DB 61 DHLTMEARAWMGIMVLEWGPQ---LLATFPGSGSALHPSPSLI 101

RESULT 13

ID AD119840 standard; protein; 101 AA.

AC AD119840;

DT 22-APR-2004 (first entry)

XX Human NOV4 protein fragment.

XX Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 KW infection; anorexia; cancer; cardiovascular disease; hypertension;
 KW Parkinson's disease; neurodegenerative disorder; Alzheimer's disease;
 KW haematopoietic disorder; inflammatory skin disorder; asthma;
 KW dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 KW haematopoiesis; wound healing; angiogenesis; chromosome mapping;
 KW tissue typing; preventive medicine; pharmacogenomic; gene therapy;
 KW anorectic; cardiant; virucide; antibacterial; fungicide; protozoacide;
 KW nootropic; neuroprotective; dermatological; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 27 /note= "Encoded by CGC"

FT Misc-difference 40 /note= "Encoded by GGG"

FT Misc-difference 51 /note= "Encoded by CTG"

XX US2004002134-A1.

XX 01-JAN-2004.

PF 15-OCT-2001; 2001US-00977819.
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201386P.
 PR 31-MAY-2000; 2000US-00584411.
 PA (CURA-) CURAGEN CORP.

PI Shimetsu RA, Fernandes ER, Herrman JL, Vernet CM;
 DR WPI; 2004-070737/07.
 DR N-PSDB; AD119839.

PT New NOVX nucleic acids encoding human KIA0768 protein-like and human
 PT protein PRO-228 polypeptides, useful for treating NOVX-associated
 PT disorders.

XX Example 3; SEQ ID NO 76; 95PP; English.

CC The present invention is based in part on the discovery of novel secreted
 CC and membrane-bound polypeptides and their encoding polynucleotides. The
 CC nucleic acids and polypeptides are collectively referred as NOVX. The
 CC invention is useful for treating, preventing and diagnosing diseases such
 CC as metabolic disorders, diabetes, obesity, infectious diseases such as
 CC viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
 CC cancer, cardiovascular diseases such as hypertension and atherosclerosis,
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders such as osteoarthritis, haematopoietic
 CC disorders, inflammatory skin disorders, asthma and various dyslipidemias.
 CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV protein
 CC fragment.

XX Sequence 101 AA;

Query Match 34.9%; Score 296; DB 8; Length 101;
 Best Local Similarity 58.1%; Pred. No. 3.1e-24;
 Matches 61; Conservative 7; Mismatches 29; Indels 8; Gaps 2;

QY 1 MPEHYIGVFPFIQVREKIHETPAVRLIGPGRKAKMGWGMGMSFSLPQACVXGAAPMOT 60
 DB 1 MPEHYIGVFPFIQVREKIHETPAVRLIGPGRKAKMGWGMGMSFSLPQACVXGAAPMOT 60

QY 61 GSENRDGSQGMDS---ALGPRGWTTPASCHLPRLQSAALHPSPSLI 101
 DB 61 DHLTMEARAWMGIMVLEWGPQ---LLATFPGSGSALHPSPSLI 101

RESULT 14

ID ABB05728 standard; protein; 703 AA.

XX ABB05728;

XX 30-APR-2002 (first entry)

XX Human transmembrane protein clone tes3_7n12.

XX Human; foetal brain; foetal kidney; melanoma; testis; amygdala;

XX gene therapy.

XX Homo sapiens.

XX WO200198454-A2.

XX 27-DEC-2001.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 15:44:07 ; Search time 24 Seconds

(without alignments)
420.014 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPELYDGVPIQLVREKH.....LSQIRPTISIFINGTSPSH 152

Scoring table: BLAST62
Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTCUS COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	10.4	1745	US-09-795-061-4	Sequence 4, Appl 1
2	84.5	10.0	334	US-09-252-991A-16629	Sequence 16629, A
3	82.5	9.7	884	US-08-463-976A-2	Sequence 2, Appl 1
4	82.5	9.7	884	US-08-982-412-2	Sequence 2, Appl 1
5	81.5	9.6	457	US-09-252-991A-29671	Sequence 29671, A
6	81	9.6	1739	US-09-795-061-2	Sequence 2, Appl 1
7	80	9.4	185	US-08-463-911-3	Sequence 3, Appl 1
8	80	9.4	365	US-09-489-847-322	Sequence 322, App
9	80	9.4	542	US-09-489-847-323	Sequence 323, App
10	79	9.3	215	US-09-511-176B-51	Sequence 51, Appl
11	79	9.3	215	US-09-619-740-54	Sequence 54, Appl
12	79	9.3	236	US-09-140-804-6	Sequence 6, Appl 1
13	79	9.3	236	US-09-686-838B-6	Sequence 6, Appl 1
14	79	9.3	272	US-09-252-991A-27652	Sequence 27652, A
15	78	9.2	75	US-09-513-999C-5144	Sequence 5144, Ap
16	77.5	9.1	178	US-09-252-991A-23713	Sequence 23713, A
17	77.5	9.1	305	US-08-712-948-2	Sequence 2, Appl 1
18	77.5	9.1	405	US-08-755-235-4	Sequence 4, Appl 1
19	77	9.1	318	US-08-633-148-4	Sequence 4, Appl 1
20	77	9.1	332	US-09-062-365-1	Sequence 1, Appl 1
21	77	9.1	340	US-08-633-148-2	Sequence 2, Appl 1
22	77	9.1	404	US-09-638-649-3	Sequence 3, Appl 1
23	77	9.1	1086	US-09-252-991A-25051	Sequence 25051, A
24	76	9.0	333	US-08-712-948-1	Sequence 1, Appl 1
25	76	9.0	772	US-09-404-879A-388	Sequence 388, App
26	76	9.0	772	US-09-667-857-388	Sequence 388, App
27	76	9.0	914	US-09-404-879A-312	Sequence 312, App

28	76	9.0	914	US-09-338-933-312	Sequence 312, App
29	76	9.0	914	US-09-667-857-312	Sequence 312, App
30	75.5	8.9	498	PCR-US94-01101-2	Sequence 2, Appl 1
31	75	8.8	593	US-09-252-991A-20441	Sequence 20441, A
32	75	8.8	1442	US-08-316-650-12	Sequence 12, Appl 1
33	75	8.8	1442	PCR-US95-02251-12	Sequence 12, Appl 1
34	74	8.7	391	US-09-270-767-58641	Sequence 58641, A
35	74	8.7	608	US-09-270-767-43297	Sequence 43297, A
36	74	8.7	1184	US-08-446-036B-20	Sequence 20, Appl 1
37	74	8.7	1184	US-08-446-010B-20	Sequence 20, Appl 1
38	74	8.7	1184	US-08-805-445-20	Sequence 20, Appl 1
39	74	8.7	1184	US-08-064-067D-20	Sequence 20, Appl 1
40	74	8.7	1184	US-09-066-208-20	Sequence 20, Appl 1
41	74	8.7	1187	US-08-357-598-8	Sequence 8, Appl 1
42	74	8.7	1187	US-08-097-997A-13	Sequence 13, Appl 1
43	74	8.7	1187	US-09-003-289-8	Sequence 8, Appl 1
44	74	8.7	1187	US-08-665-574C-13	Sequence 13, Appl 1
45	74	8.7	1187	US-08-946-994-13	Sequence 13, Appl 1

ALIGNMENTS

RESULT 1
US-09-795-061-4
; Sequence 4, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; TITLE OF INVENTION: Itamurra, Yasutada
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-061-4

Query Match

Best Local Similarity 10.4%; Score 88; DB 4; Length 1745;
Best Local Similarity 24.2%; Pred. No. 1;
Matches 45; Conservative 17; Mismatches 54; Indels 70; Gaps 9;

QY	25	VRILGPPGKAKWGMGMSFSLPQACVYG-----AAPMGTGSPN	64
DB	1407	LIGLIPPEEA-----CEKGDQGLPG--VQPPGPGDPPPGPTIGLHPSPVAGPLG	1460
QY	65	RDSQGMDCALPGRWT-----PASCHLPKOSALHPSPFLYE-----SI	105
DB	1461	QKSKSPSPSGRBDTPGAPGPGAPAEHLHGRRRRFVPLPVVGGLEEVASL	1520
QY	106	GSVCVTHYRCPSTN-----QHYN-----WAPNSELQIRTTISFIN-	145
DB	1521	TSLSTLELDLRPPGTAERPGLVCHLHRNHPHLPDGEYWDPNQCA--RDSRFVCFN	1578
QY	146	---GET 148	
DB	1579	TARGET 1584	

RESULT 2

US-09-252-991A-16629

; Sequence 16629, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: AEUUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16629
LENGTH: 334
TYPE: PR
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16629

Query Match 10.0%; Score 84.5; DB 4; Length 334;
Best Local Similarity 24.4%; Pred. No. 0.3;
Matches 38; Conservative 17; Mismatches 46; Indels 55; Gaps 11;

QY 21 ETFAVLR-LGPRGKAKWG-----MGWG-----SPSLPQACVKGAPVQGTSP 63
DB 50 ETVVWGTRFGALG-GSWGPKFPGRVGVGMPGRRAFSWPRCTGAVSP----- 102
QY 64 NRDGSGQMDGALG-SPRGW---TPASCHLPLRGSALHPSPLYESIGS---VCVTT 112
DB 103 -GGRPPIRGAFQAGPPPRAMVRSPT-----SPALVAPKGAAPRIVAVTA 148
QY 113 HPRCPSTNQHVMHLPNSLS--QIRFTISIFING 146
DB 149 ---CPTGVHPTMAALQLAGQLGALQVETQG 180

RESULT 3

US-08-465-976A-2
Sequence 2, Application US/08465976A
Patent No. 5869632
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN GILFILLAN, CECCHI
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NJ
COUNTRY: US
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465, 976A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY F
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-444
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 994-1700
TELEFAX: (201) 994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 9.7%; Score 82.5; DB 2; Length 884;

Best Local Similarity 28.9%; Pred. No. 1.8;
Matches 48; Conservative 9; Mismatches 54; Indels 55; Gaps 9;

QY 13 QLVREKLEHTPA-VLRLLGPPGKAKWGMSFSLPQACVGA-----APMOTGS----- 62
DB 641 QLSREERLEAPAPVLRPLSRPS-----QECMDAPGRLEKRGSTILPRR 686
QY 63 -PNRDGSGQMDGALG-----PRGW-----TPASCH-LPLRGSALHP 96
DB 687 QPRPDYPGAGAGRGSDALDLGAPREWSLTLPERRTRDLDSPHLCPLPSGNSQGT 746
QY 97 S-PSLIYESIGSVCTTHPRCPSTNQHVMHLPNSLSQIR 137
DB 747 SCHPRGWTLCIGARTLGSWTRCTAGTPHKEKPLGHSRSCSELGRTR 792

RESULT 4

US-08-982-412-2
Sequence 2, Application US/08982412
Patent No. 5958729
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE,
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/982, 412
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PFI91PCT2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8439
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-982-412-2

Query Match 9.7%; Score 82.5; DB 2; Length 884;
Best Local Similarity 28.9%; Pred. No. 1.8;
Matches 48; Conservative 9; Mismatches 54; Indels 55; Gaps 9;

QY 13 QLVREKLEHTPA-VLRLLGPPGKAKWGMSFSLPQACVGA-----APMOTGS----- 62
DB 641 QLSREERLEAPAPVLRPLSRPS-----QECMDAPGRLEKRGSTILPRR 686
QY 63 -PNRDGSGQMDGALG-----PRGW-----TPASCH-LPLRGSALHP 96
DB 687 QPRPDYPGAGAGRGSDALDLGAPREWSLTLPERRTRDLDSPHLCPLPSGNSQGT 746
QY 97 S-PSLIYESIGSVCTTHPRCPSTNQHVMHLPNSLSQIR 137
DB 747 SCHPRGWTLCIGARTLGSWTRCTAGTPHKEKPLGHSRSCSELGRTR 792

```
RESULT 5
US-09-252-991A-29671
; Sequence 29671, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29671
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29671

Query Match.          9.6%; Score 81.5; DB 4; Length 457;
Best Local Similarity 33.0%; Pred. No. 0.96;
Matches 33; Conservative 6; Mismatches 44; Indels 17; Gaps 4;

QY 2 PHLVDGVFPQLVREKLEHETPAVLRL-----GPGKAKWGMGMSFSLPQACVYGAA 56
DB 290 PALRRGRIP--ERRPRLAKGLRLRPVGHPPASAD-----RFPAPRQPGGR 339
QY 57 PMQTGSPNDSQGMGALGPRGWTTPASCHLPLRQSAIHP 96
DB 340 PRAGAALRRGPR--MGRLAPRGRTPAARHPLRDLALHP 377

RESULT 6
US-09-795-061-2
; Sequence 2, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-795-061-2

Query Match          9.6%; Score 81; DB 4; Length 1739;
Best Local Similarity 24.2%; Pred. No. 6.2;
Matches 44; Conservative 19; Mismatches 53; Indels 66; Gaps 10;

QY 25 VLRLLGPPGKAKWGMGMSFSLPQACVYGAAAPNQ-----TSSPN 64
DB 1406 LIGLIGRPGA-----GEKDGQLRG--VQGPRLGQGPRLGPRVSLGHPQPPGVGRLG 1459
QY 65 RDGSGCMGALGPRG-----WTPASCH-LPLRQS---ALHSPSLIYESIGSVC 109
DB 1460 QKSGSGSGSLGPRGDPGPPAPPGPPGSPALVHGLRRRSVDTLLEGLEBVMASLNSLS 1519
QY 110 VTHHPYRCPSNT-----QHYH-----WLRPNSLSQIFFTISIFIN---G 146
DB 1550 LELQQLGRLPTAESPGLMCRELHRDHPHLDBGYWDIPNGCARDAF--KVFQNFYAGG 1577
QY 147 ET 148
```

```
DB 1578 ET 1579

RESULT 7
US-08-463-911-3
; Sequence 3, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-463-911-3

Query Match          9.4%; Score 80; DB 2; Length 185;
Best Local Similarity 23.5%; Pred. No. 0.44;
Matches 23; Conservative 13; Mismatches 32; Indels 30; Gaps 4;

QY 48 PQACVYGAPMOTGSMNRGSGMGALGPRGWT-----PASCH-----L 87
DB 6 PESC--NAPGQGPQGPGPPGPPGPNMGPPGJGPPGPPGVTNCHSKGTSAF 62
QY 88 PLRQSAIHPSPS-----LIYESIGSVCTTHPYRCP 118
DB 63 AVANLPLPAPGQPVIFKEALHDAQHPDLATGVFTCP 100

RESULT 8
US-09-489-847-322
; Sequence 322, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
```


EARLIER APPLICATION NUMBER: 60/095,486
EARLIER FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: 60/096,319
EARLIER FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: 60/095,454
EARLIER FILING DATE: 1998-08-06
EARLIER APPLICATION NUMBER: 60/095,455
EARLIER FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 322
LENGTH: 365
TYPE: PRT
ORGANISM: Homo sapiens
US-09-489-847-322

Query Match 9.4%; Score 80; DB 4; Length 365;
Best Local Similarity 25.9%; Pred. No. 1.1;
Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

12 IQVREKLE-----TPAVLRLLGPPGKAKMGWGMWSFSLPQACVXGAAP 57
Db 219 LQLEERLHQLSLKDTGPAGEAGPPGPGGLQGPPEPAG----- 257
QY 58 MGTGSPNRDGSQGMGALGPRG-----WTPASCHLPLRQSALHPSPSLIYESIGSV 108
Db 258 -PFGSGKXGQGEQGPFGPPGPGQGVGGAFAA---PVPQVAFSAALSLPRSPGTV 309

RESULT 9

US-09-489-847-323
Sequence 323, Application US/09489847
Patent No. 6476195
GENERAL INFORMATION:
APPLICANT: Rosen et al
TITLE OF INVENTION: 98 Human Secreted Proteins
FILE REFERENCE: P2031P1
CURRENT APPLICATION NUMBER: US/09/489,847
CURRENT FILING DATE: 2000-01-24
EARLIER APPLICATION NUMBER: PCT/US99/17130
EARLIER FILING DATE: 1999-07-29
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
EARLIER APPLICATION NUMBER: 60/095,486
EARLIER FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: 60/096,319
EARLIER FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: 60/095,454
EARLIER FILING DATE: 1998-08-06
EARLIER APPLICATION NUMBER: 60/095,455
EARLIER FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 323
LENGTH: 542
TYPE: PRT
ORGANISM: Homo sapiens
US-09-489-847-323

Query Match 9.4%; Score 80; DB 4; Length 542;
Best Local Similarity 25.9%; Pred. No. 1.8;
Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

12 IQVREKLE-----TPAVLRLLGPPGKAKMGWGMWSFSLPQACVXGAAP 57
Db 323 LQLEERLHQLSLKDTGPAGEAGPPGPGGLQGPPEPAG----- 361
QY 58 MGTGSPNRDGSQGMGALGPRG-----WTPASCHLPLRQSALHPSPSLIYESIGSV 108
Db 362 -PFGSGKXGQGEQGPFGPPGPGQGVGGAFAA---PVPQVAFSAALSLPRSPGTV 413

RESULT 10

US-09-911-176B-51
Sequence 51, Application US/09911176B
Patent No. 6518403
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: ANTIBODIES THAT BIND AN
ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG
FILE REFERENCE: 97-30D1
CURRENT APPLICATION NUMBER: US/09/911,176B
CURRENT FILING DATE: 2001-07-23
PRIOR APPLICATION NUMBER: 09/118,408
PRIOR FILING DATE: 1998-07-17
PRIOR APPLICATION NUMBER: 60/053,154
PRIOR FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 51
LENGTH: 215
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-911-176B-51

Query Match 9.3%; Score 79; DB 4; Length 215;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

30 GPPGKAKMGWGMWSFSLPQACVXGAAPMOTGSPNRDGSQGMGALGPRGWTTPASCH--- 86
Db 52 GPGG-----TPGKPGPPGNNGPPGLPGRPDPGPGMT-VNCHSKG 88
QY 87 ---LPLRQSALHPSPS-----LIVESIGSVCTTHPRCP 118
Db 89 TSAFAVAVANLPPAPSGPVITFKEALHDAQHPDLATGVFTCP 130

RESULT 11

US-09-619-740-54
Sequence 54, Application US/09619740
Patent No. 6544946
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Lasser, Gerald W.
TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION
FILE REFERENCE: 99-12C3
CURRENT APPLICATION NUMBER: US/09/619,740
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/253,604
PRIOR FILING DATE: 1999-02-19
PRIOR APPLICATION NUMBER: 09/444,794
PRIOR FILING DATE: 1999-11-22
PRIOR APPLICATION NUMBER: 09/506,855
PRIOR FILING DATE: 2000-02-17
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 54
LENGTH: 215
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-619-740-54

Query Match 9.3%; Score 79; DB 4; Length 215;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

30 GPPGKAKMGWGMWSFSLPQACVXGAAPMOTGSPNRDGSQGMGALGPRGWTTPASCH--- 86
Db 52 GPGG-----TPGKPGPPGNNGPPGLPGRPDPGPGMT-VNCHSKG 88
QY 87 ---LPLRQSALHPSPS-----LIVESIGSVCTTHPRCP 118
Db 89 TSAFAVAVANLPPAPSGPVITFKEALHDAQHPDLATGVFTCP 130

```

RESULT 12
US-09-140-804-6
; Sequence 6, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; PRIOR FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Tamias sibiricus
US-09-140-804-6

Query Match
Best Local Similarity 9.3%; Score 79; DB 3; Length 236;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

QY 30 GPPKAKWGMGMSFSLPQACVGAAPMOTGSPNRDGSQGMGALGPRGWTASCH--- 86
DB 73 GPPG-----TPGKPGPMGNGFPGLPGPPPGMT-VNCHSKG 109
QY 87 ---LPLROSALHPSPS-----LIVESIGVCVTHTPYRCP 118
DB 110 TSAFAVKANLPPAPDSQVIFKEALHDAQGHFDLATGVFTCP 151

RESULT 13
US-09-686-838B-6
; Sequence 6, Application US/09686838B
; Patent No. 6482612
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
; FILE REFERENCE: 97-49D1
; CURRENT APPLICATION NUMBER: US/09/686,838B
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/140,804
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: US 60/056,983
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Tamias sibiricus
US-09-686-838B-6

Query Match
Best Local Similarity 9.3%; Score 79; DB 4; Length 236;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

QY 30 GPPKAKWGMGMSFSLPQACVGAAPMOTGSPNRDGSQGMGALGPRGWTASCH--- 86
DB 73 GPPG-----TPGKPGPMGNGFPGLPGPPPGMT-VNCHSKG 109
QY 87 ---LPLROSALHPSPS-----LIVESIGVCVTHTPYRCP 118
DB 110 TSAFAVKANLPPAPDSQVIFKEALHDAQGHFDLATGVFTCP 151

```

```

; Sequence 27852, Application US/09252591A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27852
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27852

Query Match
Best Local Similarity 9.3%; Score 79; DB 4; Length 272;
Matches 28; Conservative 11; Mismatches 47; Indels 18; Gaps 5;

QY 60 TSPNRDGSQGMGALGPRGWTASCHLPU-----ROSALHPSPS-----LIVESIG-- 106
DB 61 SSSRSRDSASNSGAPWPEWTPTTSSSSAPARACRRNCPHPPSRNSPAFRSIDQ 120
QY 107 SVCVTHTPYRCPSTNQHMYLPPNSELGQRTISLIFINGETSP 150
DB 121 TYCESRPPRRPGTADHRW-----RMAQFARLRRAPAPADP 159

RESULT 15
US-09-513-999C-5144
; Sequence 5144, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; PATENT REFERENCE: 59,US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5144
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: UNSURE
; LOCATION: 68
; OTHER INFORMATION: Xaa= * or Gly
US-09-513-999C-5144

Query Match
Best Local Similarity 9.2%; Score 78; DB 4; Length 75;
Matches 17; Conservative 7; Mismatches 18; Indels 8; Gaps 3;

QY 40 GWMGMSFSLPQACVGAAPMOT--GSPNRDGSQGMGALGPRGWT--ASC 85
DB 29 GWMGRIQSPQCPGSPGSPLONPORSVHKGKRGOP----PKCWXTSYAAC 74

```

Search completed: October 24, 2004, 16:04:22
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 16:00:27 ; Search time 70 Seconds

(without alignments)
703.019 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MFLHYIDGVFPIQLVREKLH.....LSQIRPTISIFINGETSPSH 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*

5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*

7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*

9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*

17: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*

19: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*

20: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	846	99.8	152 10	US-09-977-418-8
2	846	99.8	152 10	US-09-977-033A-8
3	846	99.8	152 10	US-09-977-751C-8
4	846	99.8	152 10	US-09-977-639A-8
5	846	99.8	152 11	US-09-977-819B-8
6	312	36.8	101 10	US-09-977-418-76
7	312	36.8	101 10	US-09-977-033A-76
8	312	36.8	101 10	US-09-977-751C-76
9	312	36.8	101 10	US-09-977-639A-76
10	312	36.8	101 11	US-09-977-819B-76
11	95	11.2	247 14	US-10-104-047-3142
12	92	10.8	210 16	US-10-437-963-190506
13	90.5	10.7	502 14	US-10-029-386-32930

14	88.5	10.4	92	16	US-10-437-963-113445	Sequence 113445,
15	88.5	10.4	214	16	US-10-437-963-137051	Sequence 137051,
16	88	10.4	1745	10	US-09-795-061-4	Sequence 4, Appl1
17	86	10.1	306	14	US-10-156-761-7750	Sequence 7750, Ap
18	85	10.0	1106	14	US-10-369-493-5709	Sequence 5709, Ap
19	84.5	10.0	232	15	US-10-264-049-2379	Sequence 2379, Ap
20	84	9.9	138	15	US-10-424-599-183864	Sequence 183864,
21	83.5	9.8	631	15	US-10-433-544-12	Sequence 12, Appl
22	82.5	9.7	68	14	US-10-106-598-5282	Sequence 5282, Ap
23	81.5	9.6	1279	13	US-10-087-192-1455	Sequence 1455, Ap
24	81	9.6	1739	10	US-09-795-061-2	Sequence 2, Appl1
25	80.5	9.5	1554	15	US-10-276-774-2260	Sequence 2260, Ap
26	80.5	9.5	3298	14	US-10-149-819-21	Sequence 21, Appl
27	80.5	9.5	3312	14	US-10-225-678A-656	Sequence 656, App
28	80.5	9.5	3312	15	US-10-038-854-67	Sequence 67, Appl
29	80.5	9.5	3312	16	US-10-408-765A-1499	Sequence 1499, Ap
30	80.5	9.5	4115	15	US-10-038-854-4	Sequence 4, Appl1
31	80	9.4	361	14	US-10-411-120-68	Sequence 68, Appl
32	80	9.4	365	15	US-10-351-334-322	Sequence 32, App
33	80	9.4	542	14	US-10-411-120-86	Sequence 86, Appl
34	80	9.4	542	15	US-10-351-334-323	Sequence 33, App
35	80	9.4	583	16	US-10-408-765A-1900	Sequence 1900, Ap
36	80	9.4	920	14	US-10-236-055A-26	Sequence 26, Appl
37	80	9.4	1016	16	US-10-437-963-159554	Sequence 15954, Ap
38	79.5	9.4	132	16	US-10-437-963-159554	Sequence 15954,
39	79	9.3	215	9	US-09-911-176B-51	Sequence 51, Appl
40	79	9.3	215	14	US-10-160-762-54	Sequence 54, Appl
41	79	9.3	215	14	US-10-222-857-1	Sequence 1, Appl1
42	79	9.3	215	14	US-10-360-186-54	Sequence 54, Appl
43	79	9.3	236	14	US-10-197-293-6	Sequence 6, Appl1
44	78.5	9.3	268	14	US-10-185-425-6	Sequence 6, Appl1
45	78.5	9.3	402	14	US-10-343-953-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-09-977-418-8

Sequence 8, Application US/09977418

Publication No. US20030027158A1

GENERAL INFORMATION:

APPLICANT: Shimkets et al

TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded t

FILE REFERENCES: 15966-552

CURRENT APPLICATION NUMBER: US/09/977,418

CURRENT FILING DATE: 2001-10-15

PRIOR APPLICATION NUMBER: 09/584,411

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: USSN 60/189,810

PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: USSN 60/191,158

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: USSN 60/193,086

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: USSN 60/201,388

PRIOR FILING DATE: 2000-05-03

NUMBER OF SEQ ID NOS: 93

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 152

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: Variant

LOCATION: (1)-(152)

OTHER INFORMATION: where Xaa can any amino acid

US-09-977-418-8

Query Match 99.8%; Score 846; DB 10; Length 152;

Best Local Similarity 100.0%; Pred. No. 1e-73;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152

RESULT 2
US-09-977-033A-8
Sequence 8, Application US/09977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandez, Elma
APPLICANT: Hernandez, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030082554A1el nucleic acid sequences encoding human KIAA
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977,033A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: variation
LOCATION: (53)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined by the specification
US-09-977-033A-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152

RESULT 3
US-09-977-751C-8
Sequence 8, Application US/09977751C
Publication No. US20030134430A1

GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandez, Elma
APPLICANT: Hernandez, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhabdit
TITLE OF INVENTION: Polypeptides.
FILE REFERENCE: 15966-552 CON S-40
CURRENT APPLICATION NUMBER: US/09/977,751C
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Variant
LOCATION: (53)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the specification.
US-09-977-751C-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSLSQIRFTISIFINGETSPSH 152

RESULT 4
US-09-977-639A-8
Sequence 8, Application US/09977639A
Publication No. US20030199103A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandez, Elma
APPLICANT: Hernandez, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030199103A1el amino acid sequences for human epidermal gr
TITLE OF INVENTION: polypeptides.
FILE REFERENCE: 15966-552 CON-S34
CURRENT APPLICATION NUMBER: US/09/977,639A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03/584,411
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Variant
LOCATION: (53)...(53)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the specification.
US-09-977-639A-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQMDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTVTHPRCPSEN 120
DB 61 GSPNRDGSQMDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTVTHPRCPSEN 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 5
US-09-977-819B-8
Sequence 8, Application US/09977819B
Publication No. US20040002134A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herriman, John
TITLE OF INVENTION: Vermet, Corine
TITLE OF INVENTION: Protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-826
CURRENT FILING DATE: 2001-10-15
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (53)...(53)
OTHER INFORMATION: "Xaa" = "Ile", "Leu", "Val" or "Phe"
US-09-977-819B-8
Query Match 99.8%; Score 846; DB 11; Length 152;

Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQMDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTVTHPRCPSEN 120
DB 61 GSPNRDGSQMDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTVTHPRCPSEN 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 6
US-09-977-418-76
Sequence 76, Application US/09977418
Publication No. US20030027158A1
GENERAL INFORMATION:
APPLICANT: Shimkets et al
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded t
FILE REFERENCE: 15966-552
CURRENT FILING DATE: 2001-10-15
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: USSN 60/201,388
NUMBER OF SEQ ID NOS: 93
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Homo sapiens
US-09-977-418-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQMDG---ALGPRGWTASCHLPRLRSALHPSPSLI 101
DB 61 DHLEWEARAWMGLWLEVGPO---LLATFSGSALHPSPSLI 101

RESULT 7
US-09-977-033A-76
Sequence 76, Application US/09977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Herriman, John
TITLE OF INVENTION: Vermet, Corine
TITLE OF INVENTION: Protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-824
CURRENT FILING DATE: 2001-10-15
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/137,322

PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-033A-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSQGMDC---ALGPRGWTPTASCHLPKQSLHPSPLI 101
DB 61 DHLTMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 8
US-09-977-751C-76

Sequence 76, Application US/09977751C
Publication No. US20030134430A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Hernandez, Elma
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030134430A1e1 Amino Acid Sequences for Human Caenothabdit
FILE REFERENCE: 15966-552 CON S-40
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-751C-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSQGMDC---ALGPRGWTPTASCHLPKQSLHPSPLI 101
DB 61 DHLTMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 9
US-09-977-639A-76

Sequence 76, Application US/09977639A
Publication No. US20030199103A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Hernandez, Elma
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030199103A1e1 amino acid sequences for human epidermal gr
FILE REFERENCE: 15966-552 CON S-34
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: US/09/977,639A
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-639A-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVPPILQVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSQGMDC---ALGPRGWTPTASCHLPKQSLHPSPLI 101
DB 61 DHLTMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 10
US-09-977-819B-76

Sequence 76, Application US/09977819B
Publication No. US20040002134A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Hernandez, Elma
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20040002134A1e1 nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON S-26
CURRENT APPLICATION NUMBER: US/09/977,819B


```

; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,332
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-8198-76

```

```

Query Match          36.8%; Score 312; DB 11; Length 101;
Best Local Similarity 60.0%; Pred. No. 2,1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

```

```

CY 1 MPEHYIDGVPIQVREKHEHTETPAVRLRLLGPPGAKMGWGSFSLPQACVXGAPMOT 60
    |||||
DB 1 MPEHYIDGVPIQVREKHEHTETPAVRLRLLGPPGAKMGWGSFSLPQACVXGAPMOT 60
    |||||
CY 61 GSPNRDGSQMDG---ALGPRGWTTPASCHLPLRQSALHPSPSLI 101
    |||||
DB 61 DHEEMERAMMGLWLEVGFPQ---LLATPSSQSALHPSPSLI 101
    |||||

```

```

RESULT 11
US-10-104-047-3142
; Sequence 3142; Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3142
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3142

```

```

Query Match          11.2%; Score 95; DB 14; Length 247;
Best Local Similarity 25.9%; Pred. No. 0.49;
Matches 35; Conservative 11; Mismatches 23; Indels 60; Gaps 7;

```

```

CY 2 PHLYIDGVPIQVREKHEHTETPAVRLRLLGPPGAKMGWGSFSLPQACVXGAPMOT 37
    |||||
DB 105 PHRYVAHLVRAQLTLEQLHAEVISAHSRDLGLEVGPGRPPRGIALAFWFAPMGNT 164
    |||||
CY 38 -----GMWGSFSLPQACVXGAPMOTGSPNR 65
    |||||
DB 165 SPEGSLAGLGRVSLSPAMGWGEGSGAGRPKEGKGMGALCLPGCTVTSALLPAINP-- 222
    |||||
CY 66 DSGQMDGALGPRGM 80
    |||||
DB 223 -GSGGV-GSVGRKGM 235
    |||||

```

```

RESULT 12
US-10-437-963-190506
; Sequence 190506; Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190506
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_86914C.1.pap
US-10-437-963-190506

```

```

Query Match          10.8%; Score 92; DB 16; Length 210;
Best Local Similarity 26.2%; Pred. No. 0.79;
Matches 39; Conservative 14; Mismatches 50; Indels 46; Gaps 6;

```

```

CY 21 ETPAVRLRLLGPPGAKMGWGSFSLPQACVXGAPMOTGSPNRDGSQMDGALG--- 76
    |||||
DB 14 ESRSPAMMGPPGSGRGMGVCVEIARR--WGRITVSGTARVGRGMIGRLGYVAS 71
    |||||
CY 77 ---PRGWTPASCHLPLRQSA---LHPSPSLIYE-----SIGVCVTT 112
    |||||
DB 72 VSCLSGWCASSWVP-RSSAYPSVSPSSSAVDLPTGGLTGNGSSASVPHSLGLHCVL- 129
    |||||
CY 113 HPRCRPSNTOHMYLAPNSELGIRPTIS 141
    |||||
DB 130 -----LTGNAEGHAFWTV 144
    |||||

```

```

RESULT 13
US-10-029-386-32930
; Sequence 32930; Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32930
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012313.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P51523, EVALUATE 0.00e+00
US-10-029-386-32930

```

Query Match 10.7%; Score 90.5; DB 14; Length 502;
Best Local Similarity 31.1%; Pred. No. 2.9;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKMGWG-WGMSFSLP-----QACVXGAPMGTGSPNRD-----GSGMGALGPRGW 80
DB 36 GAVSGGAGWENSTEVREXGQOATLGADGQ-GGGRGLGADGGRDGA-GPRS- 92
QY 81 TPASCHLPRLQSALHPSPLIYESIGSVG-----VTTT-----PYRCPEN 120
DB 93 EPA-----DRALRPSF--LPEEPGRCGCGKARSGSYLLQHRVATGKPYTCPEC 143
QY 121 TQHYMLRPNSELSQ 135
DB 144 GKAFAM-----SSNLSQ 155

RESULT 14
US-10-437-963-113445
Sequence 113445, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437, 963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 113445
LENGTH: 92
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17231C.1.pap
US-10-437-963-113445

Query Match 10.4%; Score 88.5; DB 16; Length 92;
Best Local Similarity 29.7%; Pred. No. 0.69;
Matches 27; Conservative 14; Mismatches 15; Indels 35; Gaps 7;

QY 30 GPPGKAKMGWG-----WGMSFSLPQAC-----VXGAAPMGTGSPNRDGSQ 69
DB 2 GPP-----WGWGHSPPRHPRGWSGMNPNRQCRVSHRGRRGTGAC-VRSIDPSIDRAR 56
QY 70 GMDGALGPRGWTTPASCHLPRLQSALHPSPL 100
DB 57 G--GVISF--TPSA-----RSSAGHRSPAT 77

RESULT 15
US-10-437-963-137051
Sequence 137051, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 137051
LENGTH: 214
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_38571C.1.pap
US-10-437-963-137051

Query Match 10.4%; Score 88.5; DB 16; Length 214;
Best Local Similarity 25.6%; Pred. No. 1.8;
Matches 40; Conservative 17; Mismatches 70; Indels 29; Gaps 7;

QY 3 HLYIDG-VFPIQLVREKLETPAVLRLLGPPGKAKMGWMSFSLPQACVXGAAPMGTG 61
DB 45 HLIVSGSSFTVAARTELGHPRTWPLRLPSSRMTW-----PLQASAGGGKETE 96
QY 62 SPNRDGSQMDGALGPRGWTTPA-----SCHL-PLRQSALHPSPLIYESIGSVCTT 113
DB 97 EGGGCGGGGHSCTP--TPADMFLVSCRRPRLRTVSTAVALALALAT-TPULTSH 152
QY 114 PYRCPSTQHYMLRPNSELSQIRFTISIFINGETS 149
DB 153 ARSSAH-----PRAQLPTFAFASATIIIGCTS 180

Search completed: October 24, 2004, 16:05:37
Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 15:41:11 ; Search time 24 Seconds

(without alignments)
609.373 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPHLYIDGVPPQLVREKLM.....LSQIRFTISIFINGETSPSH 152

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR.79:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.5	11.0	650	2	T22002
2	85	10.0	1106	2	T18739
3	83	9.8	219	2	A35650
4	82	9.7	445	2	S43492
5	81	9.6	423	2	T16750
6	80.5	9.5	1364	2	T00250
7	80	9.4	992	2	T08772
8	79	9.3	215	2	C48150
9	77.5	9.1	330	2	A40855
10	77.5	9.1	544	2	S62527
11	77.5	9.1	555	2	D95377
12	77.5	9.1	584	2	I50419
13	77	9.1	68	2	S47648
14	77	9.1	290	2	B88638
15	77	9.1	404	1	I51596
16	77	9.1	1120	2	H88449
17	76	9.1	1690	1	CGH018
18	76	9.0	1419	2	A41182
19	76	9.0	1486	1	B40333
20	76	9.0	1487	2	B41182
21	75.5	8.9	325	2	S02170
22	75.5	8.9	337	2	T23794
23	75.5	8.9	498	2	A48203
24	75.5	8.9	1744	2	S40991
25	75	8.8	460	2	T33110
26	75	8.8	1603	2	S23810
27	74.5	8.8	107	2	B51396
28	74.5	8.8	342	2	A46396
29	74.5	8.8	921	2	S40495

30	74.5	8.8	921	2	S42617	collagen alpha 1(I
31	74	8.7	313	2	T33010	hypothetical prote
32	74	8.7	1187	1	TVHUY2	proteol-tyrosine k
33	74	8.7	1747	2	A45974	collagen alpha 1(X
34	74	8.7	1857	2	S31212	collagen alpha 1(X
35	74	8.7	1888	2	S78476	collagen alpha 1(X
36	73.5	8.7	509	2	H87389	conserved hypothet
37	73.5	8.7	931	2	S13580	collagen alpha 1(I
38	73.5	8.7	1132	2	T03844	telomerase catalyt
39	73.5	8.7	1546	1	CGH02E	collagen alpha 2(X
40	73.5	8.7	1570	1	CGH018	collagen alpha 3(I
41	73.5	8.7	1591	1	S22917	collagen alpha 5(I
42	73	8.6	121	2	F72580	hypothetical prote
43	73	8.6	198	2	I49558	collagen alpha 1(I
44	73	8.6	360	2	T37285	collagen dpy-2 - C
45	73	8.6	380	2	T28868	cuticle collagen d

ALIGNMENTS

RESULT 1

T22002 hypothetical protein F39H1.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T22002

R/White, S.
submitted to the EMBL Data Library, October 1996

A/Accession: Z19500

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-650 <MIL>

A/Cross-references: UNIPROT:O17866; EMBL:Z81079; PIDN:CA803084.1; GSPDB:GN00019; CESP:F

A/Experimental source: clone F39H11

C/Genetics:

A/Map position: 1

A/Intons: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match

Best Local Similarity 38.6%; Pred. No. 0.42;

Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;

23 PAVRLRPPGKAKMGWGFSLPQACVXGAAPKQTSPPRDGSGMDALGPRG 79

DB 229 PGMPGLAGPPPGSGCTGHSAGADGPQ-----GPGGLPGAPGRDGTGVEGGRGPG 280

RESULT 2

T18739 hypothetical protein B0393.5 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T18739

R/Sulston, J.
submitted to the EMBL Data Library, September 1994

A/Accession: Z19013

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1106 <MIL>

A/Cross-references: UNIPROT:Q17494; EMBL:Z37983; NID:el519039; PIDN:CA86058.1; GSPDB:G

C/Experimental source: clone B0393

C/Genetics:

A/Map position: 3

A/Intons: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 946/1; 9

Query Match 10.0%; Score 85; DB 2; Length 1106;

Best Local Similarity 23.7%; Pred. No. 4.9;

Matches 36; Conservative 17; Mismatches 41; Indels 58; Gaps 7;

QY 12 IQLVREKLN-----ETPAVIRLL-----GPPGKAKMGWGSFSLPQACV 52
 Db 722 LSLIVKVCQARGPMTGISTPSRFIMLNNRIASGSNVACQPGK-----MMYRVYIDL 773
 QY 53 XGAAPQQTGSP-----NRDQSGMD-----GALGPPGWPAPASCH 86
 Db 774 QCTPPGGLGFLCDRCQAGHVCINCESTCHCGSVACDVTGMCPGALCRAGMGSSCD 833
 QY 87 LPLRQSLHPSPLIYESIGVCVTH-PYRC 117
 Db 834 QDIDECM-----SLVMCAVGSQCVNTRGVRC 861

RESULT 3

Sur protein - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #ext_change 09-Jul-2004
 C:Accession: A35650
 R:Dora, T.; Wang, L.H.
 Mol. Cell. Biol. 10, 4068-4079, 1990
 A:Title: An alternative non-tyrosine protein kinase product of the c-src gene in chicken
 A:Reference number: A35650; MID:90318371; PMID:2115117
 A:Accession: A35650
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-219 <DOR>
 A:Cross-references: UNIPROT:Q99370; GB:M57290; MID:9212703; PIND:AAA9076.1; PID:9212704

Query Match 9.8%; Score 83; DB 2; Length 219;
 Best Local Similarity 22.1%; Pred. No. 1.3;
 Matches 29; Conservative 15; Mismatches 39; Indels 48; Gaps 6;

QY 40 GWMGSSLPQACVXGAAPQQTGSP-----NRDQSGMD-----GALG-----PRG 79
 Db 10 GVGW-ERRPIASGSGTQLHGLGPMRRHGFVLGRDGMGRDVLGAFGMSLGERGVPA 67
 QY 80 WTPASCHLPLR-----CGALHPSPLIYESIGVCVTHPYR-----116
 Db 68 AAVALLEIMTKGLVLLQCTPRCSRCHHPSSPVSKLSTIPSTTCPCRGTHRCAGGCG 127
 QY 117 ---CPENQOHY 124
 Db 128 AVPCPAGSCHH 138

RESULT 4

S43492
 Surface antigen - hepatitis B virus (subtype adr)
 N/Alternate names: HBs antigen
 N/Contains: surface antigen pre-S1 (large envelope protein); surface antigen pre-S2 (mid
 C:Species: hepatitis B virus, HBV
 A:Variety: subtype adr
 C:Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #ext_change 09-Jul-2004
 R:Jongcarvic, L.F.; Zengraf, H.; Schroeder, C.H.
 Nucleic Acids Res. 18, 4940, 1990
 A:Title: Sequence of a replication competent hepatitis B virus genome with a prex open r
 A:Reference number: S12598; MID:90370503; PMID:2395664
 A:Accession: S43492
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-445 <JON>
 A:Cross-references: UNIPROT:Q67953; EMBL:X52939; NID:9457780; PIND:CAA37114.1; PID:94577
 A:Experimental source: subtype adr
 R:Mime, L.T.; Solomon, L.R.; Ebert, J.W.; Fields, H.
 Biochem. Biophys. Res. Commun. 195, 186-191, 1993
 A:Title: Unique prex sequence in a gibbon-derived hepatitis B virus variant.
 A:Reference number: PNO601; MID:93371402; PMID:8363598
 A:Accession: PNO601
 A:Molecule type: protein
 A:Residues: 57-95 <MIM>

A:Accession: PNO602
 A:Molecule type: protein
 A:Residues: 165-196 <M12>
 C:Genetics:
 A:Gene: S
 A:Introns: 210/3
 C:Superfamily: hepatitis B virus surface antigen
 C:Keywords: surface antigen
 F:1-445/Product: surface antigen pre-S1 (large envelope protein) #status predicted <PS1>
 F:1-164/Domain: pre-S1 domain #status predicted <PRE1>
 F:165-445/Product: surface antigen pre-S2 (middle envelope protein) #status predicted <P
 F:165-219/Domain: pre-S2 domain #status predicted <PRE2>
 F:220-445/Product: surface antigen S (small envelope protein) #status predicted <PSD>

Query Match 9.7%; Score 82; DB 2; Length 445;
 Best Local Similarity 25.0%; Pred. No. 3.5;
 Matches 40; Conservative 14; Mismatches 48; Indels 58; Gaps 9;

QY 15 VREKHEETPAVL---RLGPPGKAKMGW-----GWMGSFSLPQAC-----VXGA- 55
 Db 23 IREKLAHPHFGVGHILGNKSYMGWSSKPRGQMTNLSVPRPLGFPDPHQLDPAFGAN 82
 QY 56 -APMQGSPNRD---GSGQMDGALGP-----RGWT-----PASC 85
 Db 83 SNNPDMDFNNKQHWPEAKQVGAAGFGFTPRPHGLGWSPOAGTLTTPAAPAPAST 142
 QY 86 HLPRLQSLHPSPLIYESIGVCVTHPYRCPSNTQHY 125
 Db 143 N--RQSGROTP-----ISPLRDSHPAMQMNSTTH 173

RESULT 5

hypothetical protein R13F6.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 15-Sep-2000
 C:Accession: T16750
 R:Miller, N.
 Submitted to the EMBL Data Library, April 1994
 A:Description: The sequence of C. elegans cosmid R13F6.
 A:Reference number: Z18570
 A:Accession: T16750
 A:Status: preliminary; translated from GB/EMBL/DDBL
 A:Molecule type: DNA
 A:Residues: 1-423 <MIL>
 A:Cross-references: EMBL:U00046; NID:9470358; PID:9470367; PIND:AAC47050.1; GSPDB:GN0002
 A:Experimental source: strain Bristol N2; clone R13F6
 C:Genetics:
 A:Gene: CESP.R13F6.9
 A:Map position: 3
 A:Introns: 47/1; 77/3; 102/3; 132/2; 154/1; 182/3; 214/3; 234/2; 318/3; 345/3; 372/3; 39
 C:Superfamily: human transcription regulator MBD-4

Query Match 9.6%; Score 81; DB 2; Length 423;
 Best Local Similarity 27.0%; Pred. No. 4.1;
 Matches 31; Conservative 16; Mismatches 42; Indels 26; Gaps 5;

QY 42 GWSFSLPQACVXGAAPQQTGSPNRDQSGMDGALGPRGWTAPASCHLPLRQSLHPSPLI 101
 Db 138 GLNWSMP-----SPQIPSSPTIWOSS-----GSSIASC-----ASSPSPSYF 175

QY 102 YESIGVCVTH--PYRCPSNTQHYMKRNLSLOI--RFTISFINGETSPSH 152
 Db 176 SEDGEVQVHQRPPEHPSMAQITTFELNSRVGEVFKLVNLSITVDGTTNPSN 230

RESULT 6

MEG2 protein - human (fragment)
 T00250
 C:Species: Homo sapiens (man)
 C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #ext_change 21-Jul-2003
 C:Accession: T00250
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomic 51, 27-34, 1998
 A>Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Accession: T00250
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1364 <NAK>
 A:Cross-references: EMBL:AB011536; NID:G3449297; PIDN:BA32464.1; PID:G3449298
 A:Experimental source: brain; clone HG1044
 C:Genetics:
 A:Gene: MEGP2
 A:Map position: 3p21.2-p24.1
 F:1-28/Domain: EGF homology (fragment) <EGF>
 F:32-66/Domain: EGF homology <EGF1>
 F:124-169/Domain: laminin-type EGF-like homology <LEG>

Query Match 9.5%; Score 80.5; DB 2; Length 1364;
 Best Local Similarity 32.1%; Pred. No. 17;
 Matches 34; Conservative 4; Mismatches 33; Indels 35; Gaps 6;

QY 13 QLVREKLHETPA-VLRLLGPPGKAKMGWGSFSLPQACVXGA---APMOTGS-----62
 DB 1131 QLSERLEEARAPVLRPLRSRGS-----QECMDAAPGRLEPKRSGSTLPR 1176

QY 63 -PNDSSQGDGALG-----PRGWTPASCHLPRLQSAALHSP 98
 DB 1177 QPRDYFGAMAGRPSRDALDLAGRML-STLPPRRITDLPQP 1221

RESULT 7

T08772
 hypothetical protein DKFZps86M121.1 - human (fragment)

C:Species: Homo sapiens (man)
 C>Date: 11-Jun-1999 #sequence revision 11-Jun-1999 #text_change 09-Jul-2004
 C:Accession: T08772
 R:Ottemaender, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16474
 A:Accession: T08772
 A:Molecule type: mRNA
 A:Residues: 1-992 <OTT>
 A:Cross-references: UNIPROT:Q9Y6C2; EXBL:AL050138
 A:Experimental source: adult uterus; clone DKFZs586M121
 C:Genetics:
 A:Note: DKFZps86M121.1

A:Note: DKFZps86M121.1

Query Match 9.4%; Score 80; DB 2; Length 992;
 Best Local Similarity 25.9%; Pred. No. 13;
 Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

QY 12 IQLVREKLHE-----TPAVLRLGPPGKAKMGWGSFSLPQACVXGAAP 57
 DB 773 LQLIEDRLHQSLKDLGPPAGEAPPPGPGGPPGAG-----811

QY 58 MOTSPNDSQGDGALGPRG-----WTPASCHLPRLQSAALHSPSLYESIGSV 108
 DB 812 -PPSPGKDGEGPIGPPGQEGVEGAPAA---PVPQVAFSAALSLPSPSEPTV 863

RESULT 8

C48150
 hibernation-related protein HP-27 precursor - Siberian chipmunk

C:Species: Eutamias sibiricus (Siberian chipmunk)
 C>Date: 16-Feb-1994 #sequence revision 25-Apr-1997 #text_change 09-Jul-2004
 C:Accession: C48150; A41752
 R:Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.
 Mol. Cell. Biol. 13, 1516-1521, 1993
 A>Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like
 A:Reference number: A48150; MUID:93180798; PMID:8441393
 A:Accession: C48150
 A:Molecule type: mRNA; protein
 A:Residues: 1-215 <TKA>
 A:Cross-references: UNIPROT:Q06577; GB:DI2976; NID:g287471; PIDN:BA02353.1; PID:g287472

A:Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP27, release
 A:Note: sequence extracted from NCBI backbone (NCBIN:125948, NCBI:125949)

R:Kondo, N.; Kondo, J.
 J. Biol. Chem. 267, 473-478, 1992
 A>Title: Identification of novel blood proteins specific for mammalian hibernation.
 A:Reference number: A41752; MUID:92112696; PMID:1730610
 A:Accession: A41752

A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 31-40, 'A', 42-50, 'Q', 52-215 <KON>
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho
 C:Keywords: glycoprotein; hibernation; plasma
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-215/Product: hibernation-related protein HP-27 #status experimental <MAT>
 F:91-214/Domain: complement C1q carboxyl-terminal homology <CIQ>
 F:155/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.3%; Score 79; DB 2; Length 215;
 Best Local Similarity 23.5%; Pred. No. 3;
 Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

QY 30 GPFGKAKMGWGSFSLPQACVXGAAPMOTGSPPNDSQGDGALGPRGTPASCH---86
 DB 52 GPFG-----TGGKPPGPMNGPPGLPGPPGPMNT-VNCHSG 88

QY 87 ---LPLRQSAALHSPS-----LYESTIGSVCTTHPRCP 118
 DB 89 TSAFAVKANLPPASQPVIFKRALHDAQGHFDLATGVTCP 130

RESULT 9

A40855
 homeotic protein Hox 11 - human

N:Alternate names: tcl-3 proto-oncogene
 C:Species: Homo sapiens (man)
 C>Date: 28-Feb-1992 #sequence revision 28-Feb-1992 #text_change 09-Jul-2004
 C:Accession: A40855; S17311; A41224
 R:Hatanou, M.; Roberts, C.W.M.; Minden, M.; Crist, W.M.; Koremeyer, S.J.
 Science 253, 79-82, 1991
 A>Title: Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia.
 A:Reference number: A40855; MUID:91289163; PMID:1676542
 A:Accession: A40855

A:Molecule type: mRNA

A:Residues: 1-330 <HAT>
 A:Cross-references: UNIPROT:P31314; GB:S38742; NID:g232582; PIDN:AA19293.1; PID:g23258

R:Lu, M.; Gong, Z.; Shen, W.; Ho, A.D.
 EMBO J. 10, 2905-2910, 1991
 A>Title: The tcl-3 proto-oncogene altered by chromosomal translocation in T-cell leukem

A:Reference number: S17311; MUID:92007734; PMID:1717256

A:Accession: S17311

A:Molecule type: mRNA

A:Residues: 1-275, 'L', 277-330 <LUM>
 A:Cross-references: EMBL:M62626; NID:g339199; PIDN:AAA36719.1; PID:g387683

R:Kennedy, M.A.; Gonzalez-Sarmiento, R.; Kees, U.R.; Lampert, F.; Dear, N.; Boehm, T.;

Proc. Natl. Acad. Sci. U.S.A. 88, 8900-8904, 1991

A>Title: HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24.

A:Reference number: A41224; MUID:92020958; PMID:1681546

A:Accession: A41224

A:Molecule type: DNA

A:Residues: 1-189, 'VALSPFTVTRRT', 190-275, 'L', 277-330 <KEN>
 A:Cross-references: GB:M75952
 A:Note: the authors suggest a long form with twelve additional amino acids from possibl
 C:Genetics:
 A:Gene: GDB:HOX11
 A:Cross-references: GDB:119607; OMIM:186770
 A:Map position: 10q24-10q24
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; leukemia; nucleus; proto-oncogene; transcription reg
 F:202-258/Domain: homeobox homology <HOX>

Query Match 9.1%; Score 77.5; DB 2; Length 330;
 Best Local Similarity 27.1%; Pred. No. 6;
 Matches 32; Conservative 10; Mismatches 59; Indels 17; Gaps 4;

QY 30 GPPKAKMGWGM-----SFSLPQACVXGAAPMOTGSPNRDSCQMDGALGPRGWTAS 84
 Db 81 GPGGPGAGGAGACMGWGLTGYNNMALAGGPGGGG-----GSSGAGALSAAGVTRVP 135
 QY 85 CHPLPGSALHPSPSLVIESGVCTTHYRCPST---QHNMRLPMSLSQIRFT 139
 Db 136 AHRPLGAVAHPPPL---ATGLPTVPSPVAMPGVNLTGLTTPMWSNRRYTKCRFT 189

RESULT 10

S62527

Probable amino-acid permease - fission yeast (*Schizosaccharomyces pombe*)C/Species: *Schizosaccharomyces pombe*

C/Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004

C/Accession: T41435; T39133; S62527

R/Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1998

A/Reference number: Z21993

A/Accession: T41435

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-544 <SEE>

A/Cross-references: UNIPROT:Q09887; EMBL:AL032824; PIDN:CA837426.1; GSPDB:GN00068; SPDB:

R/Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, November 1995

A/Reference number: Z21830

A/Accession: T39133

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-544 <LYE>

A/Cross-references: EMBL:Z65659; NID:G1052533; PIDN:CAA91521.1; PID:G1052542; SPDB:SPAC8

C/Genetics: <SEE>

A/Map position: 13

C/Genetics: <LYE2>

A/Map position: 11

A/Map position: 1

C/Superfamily: choline transport protein

Query Match 9.1%; Score 77.5; DB 2; Length 544;
 Best Local Similarity 23.6%; Pred. No. 12;
 Matches 25; Conservative 12; Mismatches 54; Indels 15; Gaps 2;

QY 22 TPVALLLGPFGKXKMGWGMGFSLPQACVXGAAPMOTGSPNRDSCQMDGALGPRGWT 81
 Db 73 TPMMV-----WGLIAMVYQCVANGMAELCSMTSGCLYAAAVLAPKMG 120
 QY 82 PASCHLPLRQAL--HPSPSLVIESGVCTTHYRCPSTQHY 124
 Db 121 PFAMWLTGWSNYLVQTPSPVAXSPAGXILTLVQLHNPFETQNY 166

RESULT 11

D93777

Probable cerebroside-sulfatase (BC 3.1.6.8) [imported] - *Sinorhizobium meliloti* (strainC/Species: *Sinorhizobium meliloti*

C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004

C/Accession: D93777

R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows

Proc. Natl. Acad. Sci. U.S.A. 95, 9883-9888, 2001

A/Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*

A/Reference number: A9562; MUID:21396509; PMID:11481432

A/Accession: D93777

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1555 <KUR>

A/Cross-references: UNIPROT:Q92YF6; GB:AE006469; PIDN:AAK65582.1; PID:G14524062; GSPDB:G

A/Experimental source: strain 1021, megaplasmid pSymA

R/Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, A.P.; Ampe, F.; Barloy-Hubler,

Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, E.; Komp, C.; Lelaure,
 hebbalt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
 A/Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A/Reference number: A96059; MUID:21368234; PMID:11474104
 A/Content: annotation
 C/Genetics:
 A/Gene: Sma1683
 A/Genome: plasmid
 C/Keywords: sulfuric ester hydrolase

Query Match

9.1%; Score 77.5; DB 2; Length 555;

Best Local Similarity 35.9%; Pred. No. 12;

Matches 28; Conservative 2; Mismatches 23; Indels 25; Gaps 4;

QY 29 LGPPKAKMGWGM-----WSFSLPQACVXGAAPMOTGS-PNRDGS---68
 Db 49 LGPGY---GGRGMRMPPTNIDRLADEGMTFFSYAQSCTPGAAVQTGIRPFRSGMTTV 105
 QY 69 --QMDGALGPRGWTAS 84
 Db 106 AFQCGGGLPAAEWTLAS 123

RESULT 12

I50419

g-glycerin precursor - chicken

C/Species: *Gallus gallus* (Chicken)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004

C/Accession: I50419

R/Taira, E.; Takaha, N.; Tanitua, H.; Klm, C.H.; Miki, N.

Neuron 12, 861-872, 1994

A/Title: Molecular cloning and functional expression of glicerol, a novel cell adhesion m

A/Reference number: I50419; MUID:94213753; PMID:8161457

A/Accession: I50419

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-584 <TAI>

A/Cross-references: UNIPROT:Q09089; GB:D38559; NID:G1009246; PIDN:BA07563.1; PID:G55970

Query Match 9.1%; Score 77.5; DB 2; Length 584;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 5; Mismatches 25; Indels 17; Gaps 5;

QY 37 WGMGWSFSLPQACVX---GAA-----PMOTGSPNRDSCQMDGALGPRGWTAS 86
 Db 13 WGMGWSFSLPQACVX---GAA-----PMOTGSPNRDSCQMDGALGPRGWTAS 86
 QY 87 LPLRQAL-HPSP 98
 Db 67 LHLRRVLCQPRP 79

RESULT 13

S47648

nitric oxide synthase - bovine

C/Species: *Bos primigenius* taurus (cattle)

C/Date: 26-Dec-1994 #sequence_revision 26-May-1995 #text_change 26-May-1995

C/Accession: S47648

R/Venem, R.C.; Nishida, K.; Alexander, R.W.; Harrison, D.G.; Murphy, T.J.

Biochim. Biophys. Acta 1218, 413-420, 1994

A/Title: Organization of the bovine gene encoding the endothelial nitric oxide synthase.

A/Reference number: S47648; MUID:9435350; PMID:7519447

A/Accession: S47648

A/Status: preliminary; translation not shown

A/Molecule type: DNA

A/Residues: 1-68 <VEN>

A/Cross-references: EMBL:L27056

Query Match 9.1%; Score 77; DB 2; Length 68;
 Best Local Similarity 32.3%; Pred. No. 13;
 Matches 21; Conservative 5; Mismatches 27; Indels 12; Gaps 3;

QY 24 AYVLELPPGKAKWGMGWSFSL-----PQACVXGAAPMOTGSPNPD-----GSGQMG 73
 Db 6 SVASPPPPF--AAWGMGMAAGYAAASRAQPPRHLPAAAPPPPPRTTRQTGTGEGHGRGTG 63

QY 74 ALGPR 78
 Db 64 RLGAR 68

RESULT 14
 B88638
 protein F58F6.2 [imported] - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C/Accession: B88638
 R/Anonymous: The C. elegans Sequencing Consortium.
 S/Science 282, 2012-2018, 1998
 A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A/Reference number: A75000; MUID:99069613; PMID:9851916
 A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A/Accession: B88638
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-290 <STO>
 A/Cross-references: UNIPROT:O44173; GB:chr_IV; PID:AA88359.1; PID:g2662600; GSPDB:GNOC
 C/Genetics:
 A/Note: Similar to cuticular collagen
 A/Gene: F58F6.2
 A/Map position: 4

Query Match 9.1%; Score 77; DB 2; Length 280;
 Best Local Similarity 37.7%; Pred. No. 6.4;
 Matches 23; Conservative 1; Mismatches 19; Indels 18; Gaps 3;

QY 30 GPPGKAKWGMGWSFSLPQACVXGAAPMOTGSPNPDGSGQMGALPPRGWT--PASCHL 87
 Db 142 GPPGKPG-----PQ-----GAPGDIGQPGNPGERGMDEPDPGPTGPPGPHGL 185

QY 88 P 88
 Db 186 P 186

RESULT 15
 I61596
 advanced glycosylation end-products receptor precursor - human
 N/Alternate names: advanced glycosylation end product-binding protein, 35k; glycoprotein
 C/Species: Homo sapiens (man)
 C/Date: 24-May-1996 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
 C/Accession: I61596; B42879; S27968
 R/Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko, H.;
 Genomics 23, 408-419, 1994
 A/Title: Three genes in the human MHC class III region near the junction with the class
 II region of mouse mammary tumor gene int-3.
 A/Reference number: A55562; MUID:95137587; PMID:7835890
 A/Accession: I61596
 A/Status: nucleic acid sequence not shown; translation not shown; translated from GB/ENH
 A/Molecule type: DNA
 A/Residues: 1-404 <RES>
 A/Cross-references: UNIPROT:Q15109; GB:D8769; NID:G561657; PID:BA05958.1; PID:G561659
 R/Neepert, M.; Schmidt, A.M.; Bretz, U.; Yan, S.D.; Wang, F.; Pan, Y.C.; Elliston, K.; St
 u. Biol. Chem. 267, 14996-15004, 1992
 A/Title: Cloning and expression of a cell surface receptor for advanced glycosylation en
 A/Reference number: A42879; MUID:92340547; PMID:1378843
 A/Accession: B42879
 A/Molecule type: mRNA
 A/Residues: 'G', 2-99, 'R', 101-404 <NEE>
 A/Cross-references: EMBL:M91211; NID:g190845; PID:AAA03574.1; PID:g190846
 A/Experimental source: lung
 A/Note: Sequence extracted from NCBI backbone (NCEIP:109438)
 C/Comment: Advanced glycosylation end products are heterogeneous nonenzymatically glycos

cellular function, thus contributing to tissue lesions in diabetes.
 C/Comment: This receptor appears also to mediate the effects of amyloid beta peptide on
 ates in the neurotoxic pathway that produces dementia in Alzheimer's disease.

A/Genetics:
 A/Gene: GDB:AGER
 A/Cross-references: GDB:306354; OMIM:600214
 A/Map position: 6p21.3-6p21.3
 A/Introns: 18/1; 53/3; 119/1; 140/3; 170/1; 231/1; 274/3; 322/1; 331/1; 373/2
 C/Function:
 A/Description: neuronal receptor for amphoterin, a DNA-binding protein involved in neur
 C/Superfamily: advanced glycosylation end products receptor; immunoglobulin homology
 C/Keywords: Alzheimer's disease; glycoprotein; receptor; transmembrane protein
 F/1-22/Domain: signal sequence #status predicted <SIG>
 F/23-404/Product: advanced glycosylation end products receptor #status predicted <MAT>
 F/23-44/Domain: extracellular #status predicted <EXT>
 F/31-101/Domain: immunoglobulin homology <IM1>
 F/137-210/Domain: immunoglobulin homology <IM2>
 F/252-303/Domain: immunoglobulin homology <IM3>
 F/345-362/Domain: transmembrane #status predicted <TM>
 F/363-404/Domain: intracellular #status predicted <INT>
 F/25, 81/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/38-99, 144-208, 259-301/Disulfide bonds: #status predicted

Query Match 9.1%; Score 77; DB 1; Length 404;
 Best Local Similarity 23.3%; Pred. No. 9.3;
 Matches 41; Conservative 16; Mismatches 57; Indels 62; Gaps 7;

QY 8 GVFPQLVREKLEHPAVRLRLGPPGKAKWGMGWSFSLPQACVXGAAPM----- 59
 Db 184 GLFTLD---SELMVTPA-----RGDPRFTSCSFSPGLPHRLRLTAIPQRYWEPVP 234

QY 60 -----TGSPPNRDGSQMGALGPPRGWTPASCHLPDRS-----ALHPSPLI 101
 Db 235 LEEVQLVPEP-----GGAVALPGGIVTLTEVVAQPSQLMKMDGVPLPLPSPVLI 287

QY 102 YESIGS-----VCYTHPRCPSTNQHMYLRLPNSLSQIFTSIFNGSTSS 151
 Db 288 LPEIGPDQGTYSVATSHSHGPOESR-----AVSILIEPGSEGP 329

Search completed: October 24, 2004, 16:03:52
 Job time : 26 secs

QY 74 A-----LGPR-GWTPAS-----CHLPLRGSALHPSP-- 98
 DB 272 ARHRELESLVPLRGVYKATKPTKTPGPGFTFABLRARTGVCHRSFEAKLPFCPOC 331
 QY 99 -SLIYSIGSVCTTHPYRCPSTNTQYHW-----LRPNSELQRTFTSIFINGET 148
 DB 332 SAVLY--CGBACLRADWQRCDDVSHRFMCPRLAEMERAGELATLPFYTAELVSET 387
 RESULT 2
 Q8NA12 PRELIMINARY; PRT; 247 AA.
 ID Q8NA12
 AC Q8NA12
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ35323.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX PubMed=14702039;
 Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 Sudo H., Hoshiro T., Kaku Y., Kodaera H., Kondo H., Sugawara M.,
 Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 Nishimura K., Ishibashi T., Yamashita H., Murakami K., Fujimori K.,
 Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 Oono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 Togiya S., Komai R., Hara R., Takeuchi K., Arita M., Imose N.,
 Musashino K., Yuki H., Oshino A., Sasaki N., Aotsuka S.,
 Morioka H., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukumizu Y.,
 Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 Oono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
 Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 Togaashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL: AK092642; BAC03933.1;
 SQ SEQUENCE 247 AA; 25975 MW; D69A7CCD02DC54C1 CRC64;

DB 223 -GSGGV-GSVGRKGM 235
 RESULT 3
 ID 017866 PRELIMINARY; PRT; 650 AA.
 AC 017866;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE CLE-1C protein (C. elegans CLE-1 protein) (Corresponding sequence
 C36B1.1c).
 GN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloiderinae; Caenorhabditis.
 NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Blamaa H., Pihlajaniemi T., Kuo C.J.,
 Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 collagen affects cell migration and axon guidance.";
 RL J. Cell Biol. 152:1219-1232(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF164959; AA047825.1;
 DR EMBL: Z81079; CAB03084.1;
 DR PIR: T22002;
 DR HSSP: P39061; 1XOE.
 DR WormPep: C36B1.1C; CE17816.
 DR GO:0005737; C:cytoplasm; IEA.
 DR GO:0006817; P:phosphate transport; IEA.
 DR InterPro: IPR006160; Collagen.
 DR InterPro: IPR010515; Endostatin.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF06482; Endostatin; 1.
 KW Collagen.
 SQ SEQUENCE 650 AA; 69596 MW; 6CF29ED9C16B170E CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 650;
 Best Local Similarity 38.6%; Pred. No. 3.1;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 QY 23 PAVLRLLGPPGKXKMGWGMSFSLPQACVXGAAPMOTGSPNDRSGQMGALGPRG 79
 DB 229 PGMFGLAGPFGPGSGSGHSGSGADFPQ-----GPGFLPAGRGDTSVSGQGRPG 280
 RESULT 4
 ID 09U9K6 PRELIMINARY; PRT; 778 AA.
 AC 09U9K6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE CLE-1B protein (C. elegans CLE-1 protein) (Corresponding sequence
 C36B1.1b).
 GN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Elamra H., Pihlajaniemi T., Kuo C.J.,
 RA Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 RT collagen affects cell migration and axon guidance.";
 RL J. Cell Biol. 152:1219-1232(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Leonard N.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF164959; AAD47824.1; -;
 DR EMBL; 280215; CAD21700.1; -;
 DR EMBL; 281079; CAD21700.1; JOINED.
 DR EMBL; 281079; CAD21643.1; -;
 DR EMBL; 280215; CAD21643.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR WormPep; C36B1.1b; CE29706.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR010515; Endostatin.
 DR InterPro; IPR003129; TSP_N.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF06482; Endostatin; 1.
 DR SMART; SM00210; TSPN; 1.
 DR Collagen.
 KW SEQUENCE.
 SQ 778 AA; 83787 MW; 743A7FCA7FF1ABD CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 778;
 Best Local Similarity 38.6%; Pred. No. 3.7;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 Oy 23 PAVRLLLGPPGKAKWGWGWSFSLPQACVGAAPMOTGSPNRDGSQMGALGPRG 79
 Db 357 FGMPLGAPGPPGPGSGTGHGSGADGPQ-----GPPGLGAPGARDGTSGVEGGRGPG 408
 RESULT 5
 ID Q909K7 PRELIMINARY; PRT; 1117 AA.
 AC Q909K7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DT C1F-1A protein (C. elegans C1F-1 protein) (Corresponding sequence
 DE C36B1.1a); ORFNames=C36B1.1;
 GN Name=C1F-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxId=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Elamra H., Pihlajaniemi T., Kuo C.J.,
 RA Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 RT collagen affects cell migration and axon guidance.";
 RL J. Cell Biol. 152:1219-1232(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Leonard N.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF164959; AAD47823.1; -;
 DR EMBL; 280215; CAD21701.1; -;
 DR EMBL; 281079; CAD21701.1; JOINED.
 DR EMBL; 281079; CAD21644.1; -;
 DR EMBL; 280215; CAD21644.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR WormPep; C36B1.1a; CE29705.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR010515; Endostatin.
 DR InterPro; IPR003961; FN_III-like.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003129; TSP_N.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF06482; Endostatin; 1.
 DR Pfam; PF00041; FN3; 2.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR Collagen.
 KW SEQUENCE.
 SQ 1117 AA; 121680 MW; CF323D42E88770AD CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 1117;
 Best Local Similarity 38.6%; Pred. No. 5.6;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 Oy 23 PAVRLLLGPPGKAKWGWGWSFSLPQACVGAAPMOTGSPNRDGSQMGALGPRG 79
 Db 696 FGMPLGAPGPPGPGSGTGHGSGADGPQ-----GPPGLGAPGARDGTSGVEGGRGPG 747
 RESULT 6
 ID Q810G6 PRELIMINARY; PRT; 1154 AA.
 AC Q810G6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DT C. elegans C1F-1 protein (Corresponding sequence C36B1.1d).
 DE Name=C1F-1; ORFNames=C36B1.1;
 GN Name=C1F-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxId=6239;

```

OK NCBI_TaxID=6239;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
EX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RT Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lemard N.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA White S.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z80215; CAD56563.1; -.
DR EMBL; Z81079; CAD56563.1; JOINED.
DR EMBL; Z81079; CAD56563.1; -.
DR EMBL; Z80215; CAD56563.1; JOINED.
DR HSSP; P39061; 1KOE.
DR WormBep; C36B1.1d; CE32319.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005198; Structural molecule activity; IEA.
DR GO; GO:0007155; P-cell adhesion; IEA.
DR GO; GO:0006817; Phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008995; Coma like lec_g1.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003129; TSP_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF06482; Endostatin; 1.
DR Pfam; PF00041; fn3; 2.
DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00853; FN3; 2.
KW Collagen.
SQ SEQUENCE 1154 AA; 125924 MW; 7B2D9B9B3CD1ABE5 CRC64;

Query Match 11.0%; Score 93.5; DB 2; Length 1154;
Best Local Similarity 38.6%; Pred. No. 5.8;
Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;

QY 23 PAVLRLLGPPGKAKMGWGSFSLPQACVGAAPMGTGSPVRDSSQGMGALGPGRG 79
Db 733 PCMPGLAGPPPPGSGTGHGSGADGPQ-----GPRGLGAPGRDSTGVGGGPGPG 784

RESULT 7
Q6ZNTS PRELIMINARY; PRT; 498 AA.
AC Q6ZNTS;
DT 05-JUL-2004 (TREMREL. 27, Created)
DT 05-JUL-2004 (TREMREL. 27, Last sequence update)
DT 05-JUL-2004 (TREMREL. 27, Last annotation update)
DE Hypothetical protein FLJ16083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hitaoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
```

```

RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Nagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131209; BADI8400.1; -.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 14.
DR ProDom; PD000003; ZnF_C2H2; 12.
DR SMART; SM0355; ZnF_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
SQ SEQUENCE 498 AA; 54712 MW; 7C662E7C22946B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKMGW-WGWSFSLP-----QACVGAAPMGTGSPNRD--GSGMDGALGPGRW 80
Db 32 GAVSGGAWGAMENSTVEPRAGDQGRQATLGAADBPQ-GGPRGLGPADGGRDGA-GPRS- 88

QY 81 TPASCHLPLRQSALHPSPSLIVSISGVC-----VTTT-----PYRCPSN 120
Db 89 EPA-----DRALRPPD--LPBPGRGCGCGKASQGSYLQHRVRHTGKPYTCPEC 139

QY 121 TQHYMLRPNSELSQ 135
Db 140 GKAPAW---SSNLSQ 151

RESULT 8
Q6ZTD2 PRELIMINARY; PRT; 498 AA.
AC Q6ZTD2;
DT 05-JUL-2004 (TREMREL. 27, Created)
DT 05-JUL-2004 (TREMREL. 27, Last sequence update)
DT 05-JUL-2004 (TREMREL. 27, Last annotation update)
DE Hypothetical protein FLJ44773.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Cerebellum;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Mitsuhashi K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK126727; BAC86660.1; -.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 14.
DR ProDom; PD000003; ZnF_C2H2; 11.
DR SMART; SM0355; ZnF_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
SQ SEQUENCE 498 AA; 54652 MW; 7C762F70538B1B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKMGW-WGWSFSLP-----QACVGAAPMGTGSPNRD--GSGMDGALGPGRW 80
Db 32 GAVSGGAWGAMENSTVEPRAGDQGRQATLGAADBPQ-GGPRGLGPADGGRDGA-GPRS- 88

QY 81 TPASCHLPLRQSALHPSPSLIVSISGVC-----VTTT-----PYRCPSN 120
```

```

Db      89 EPA-----DRALRSP--LPBPGRCGCGCAKAFSGSYLLQHRRVHTGKPYTCPEC 139
Qy      121 TOHYHWLRPNSELQ 135
       : : : : :
Db      140 GKAFAM---SSNLSQ 151

RESULT 9
BAC6660 PRELIMINARY; PRT; 498 AA.
AC BAC6660:
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE CDNA FLJ44773 fis, clone BRACE3034183, moderately similar to zinc
DE finger protein 84.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Arita M., Mueashiro K., Yuuki A., Tanase T., Imose N., Takeuchi K.,
RA Oshima A., Takahashi-Fujii A., Tanase T., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Salto K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuna K.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T.,
RA "NEDO human cDNA sequencing project."
RT Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK16727; BAC6660.1; -.
SQ SEQUENCE 498 AA; 5452 MW; 7C762F7053881B0F CRC64;

Query Match
Best Local Similarity 31.1%; Score 90.5; DB 2; Length 498;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

Qy      33 GKAKWGCG-WGWSFSLP-----QACVYGAAPMOTGSPNRD---GSGQMDGALGPRGM 80
Db      32 GAVSGCGWAMENSTVEPRAGDQGRQATLGAADQ--GGPERELPADGGRGA-GPRS- 88
Qy      81 TPASCHLPLRQALHPSPSLYESIGSYVC-----VTH-----PYRCPN 120
Db      89 EPA-----DRALRSP--LPBPGRCGCGCAKAFSGSYLLQHRRVHTGKPYTCPEC 139
Qy      121 TOHYHWLRPNSELQ 135
Db      140 GKAFAM---SSNLSQ 151

RESULT 10
BAD18400 PRELIMINARY; PRT; 498 AA.
AC BAD18400:
DT 12-MAY-2004 (TREMBlrel. 27, Created)
DT 12-MAY-2004 (TREMBlrel. 27, Last sequence update)
DE CDNA FLJ16083 fis, clone NT2R12008952, moderately similar to ZINC
DE FINGER PROTEIN 84.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Ishida T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotta T., Hitaoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Salto K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

```

```

RA Wagatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131209; BAD18400.1; -.
SQ SEQUENCE 498 AA; 54712 MW; 7C662E7C22946B0F CRC64;

Query Match
Best Local Similarity 31.1%; Score 90.5; DB 2; Length 498;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

Qy      33 GKAKWGCG-WGWSFSLP-----QACVYGAAPMOTGSPNRD---GSGQMDGALGPRGM 80
Db      32 GAVSGCGWAMENSTVEPRAGDQGRQATLGAADQ--GGPERELPADGGRGA-GPRS- 88
Qy      81 TPASCHLPLRQALHPSPSLYESIGSYVC-----VTH-----PYRCPN 120
Db      89 EPA-----DRALRSP--LPBPGRCGCGCAKAFSGSYLLQHRRVHTGKPYTCPEC 139
Qy      121 TOHYHWLRPNSELQ 135
Db      140 GKAFAM---SSNLSQ 151

RESULT 11
Q9J104 PRELIMINARY; PRT; 1737 AA.
AC Q9J104:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Alpha 4 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=20428740; PubMed=10852920;
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.,
RT "Schwann cells synthesize type V collagen that contains a novel alpha
RT 4 chain. Molecular cloning, biochemical characterization, and high
RT affinity heparin binding of alpha 4(V) collagen."
RL J. Biol. Chem. 275:28208-28215(2000).
DR EMBL; AF272651; AAF76432.1; -.
DR HSSP; P04278; 1D25.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; CnA_like_1ec_g1.
DR InterPro; IPR000885; F1D_collagen_C.
DR InterPro; IPR003129; TSP_N.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF02210; Laminin_G_2; 1.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD002078; F1D_collagen_C; 1.
DR Collagen.
SQ SEQUENCE 1737 AA; 171574 MW; D635D57481C257 CRC64;

Query Match
Best Local Similarity 25.3%; Score 88; DB 2; Length 1737;
Matches 46; Conservative 19; Mismatches 51; Indels 66; Gaps 11;

Qy      25 VRLILGPPGAKKWKWGKWSFSLPQACVYGAAPMOTGSPNRD---GSGQMDGALGPRGM 64
Db      1404 LIGLIGPPGPA---GEKDGQJFG--VGGPGLQDPLGPGVSGLGHGPPGVVPLIG 1457

```

QY 65 RDGSGMDGALGPRG-----WTPASCH-LPFRGSALH-PPSPSL--IYESIGSYC 109
 Db 1458 QKSGKSGPSPGSLGPRGDPGAPGPPGPPSPGPAEVLGRLRRRSLVLTPEGGLEEYVASTLS 1517
 QY 110 VTHPRYCPENT-----QHYH-----WLRPNSLSQIRFTISIFIN-----G 146
 Db 1518 LELHQLRPPGTABSPGLICHELHRNHPHLPDGEYWDPNQCARDAF--KVFCHPTAGG 1575
 QY 147 ET 148
 Db 1576 ET 1577

RESULT 12
 ID CA35_HUMAN STANDARD; PRT; 1745 AA.
 AC P25940; Q9NZG6;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Collagen alpha 3 (V) chain precursor.
 GN Name=COL5A3;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart, and Placenta;
 RX MEDLINE=20187594; PubMed=10722718;
 RA Imanura Y., Scott I.C., Greenspan D.S.;
 RT "The pro-alpha 3 (V) collagen chain. Complete primary structure,
 RT expression domains in adult and developing tissues, and comparison to
 RT the structures and expression domains of the other types V and XI
 RT procollagen chains.";
 RL J. Biol. Chem. 275:8749-8759(2000).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 479-564; 665-709; 723-758; 787-816; 922-1008;
 RP 1054-1088; 1248-1287 AND 1313-1334.
 RC TISSUE=Placenta;
 RX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 RT characterization by partial sequencing.";
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
 CC -1- FUNCTION: Type V collagen is a member of group I collagen
 CC (fibrillar forming collagen). It is a minor connective tissue
 CC component of nearly ubiquitous distribution. Type V collagen binds
 CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
 CC -1- SUBUNIT: Trimers of two alpha 1 (V) and one alpha 2 (V) chains in
 CC most tissues and trimers of one alpha 1 (V), one alpha 2 (V), and
 CC one alpha 3 (V) chains in placenta.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF177941; AAF59902.1; -
 CC PIR: S20375; S20375
 CC DR GeneW: HGNC:14864; COL5A3.
 CC DR MIM: 120216; -
 CC DR GO: GO:0005588; C:collagen type V; TAS.
 CC DR GO: GO:0005201; F:extracellular matrix structural constituent; NAS.
 CC DR InterPro: IPR008161; C1g helix.
 CC DR InterPro: IPR008160; Collagen.

DR InterPro: IPR008985; CONA like lec g1.
 DR InterPro: IPR008985; Fib collagen_C.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR003129; TSP_N.
 DR Pfam: PF01410; COLFI_1.
 DR Pfam: PF01391; Collagen; 17.
 DR Pfam: PF02210; TSP_N; 1.
 DR Prodom: PD000007; C1g helix; 2.
 DR Prodom: PD002078; Fib collagen_C; 1.
 DR SMART: SM00038; COLFI_1.
 DR SMART: SM00282; LamC; 1.
 DR SMART: SM00210; TSPN; 1.
 DR Collagen; Connective tissue; Direct protein sequencing;
 KW Extracellular matrix; Hydroxylation; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 1745 Collagen alpha 3 (V) chain.
 FT DOMAIN 31 210 TSP N-terminal.
 FT DOMAIN 211 381 Nonhelical region.
 FT DOMAIN 392 1489 Triple-helical region.
 FT DOMAIN 1490 1745 Nonhelical region.
 SQ SEQUENCE 1745 AA; 172051 MW; 5B8FF97135397AC1 CRC64;

Query Match 10.4%; Score 88; DB 1; Length 1745;
 Best local similarity 24.2%; Pred. No. 30;
 Matches 45; Conservative 17; Mismatches 54; Indels 70; Gaps 9;

QY 25 VLRLGPPGAKKMGWNGMSFSLPQACVXG-----AAPMGTGSPN 64
 Db 1407 LIGLIGPPGEX-----GEKDGQLFG--VQPPGPKGDPGPPGSLGHPGPPGAVPLG 1460
 QY 65 RDGSGMDGALGPRGWI-----PASCHLPFRGSALHSPSLIV-----SI 105
 Db 1461 QKSGKSGPSPGSLGPRGDPGAPGPPGPPSPGPAEVLGRLRRRFPVPLPVEGSGLEEVASTLS 1520
 QY 106 GSVCTVTHPRYCPSENT-----QHYH-----WLRPNSLSQIRFTISIFIN- 145
 Db 1521 TSLSLLEQLRRPPTABSPGLVCHLHRNHPHLPDGEYWDPNQCARDAF--RDSFRVFCNR 1578
 QY 146 ---GET 148
 Db 1579 TAGGET 1584

RESULT 13
 ID Q9H2C6 PRELIMINARY; PRT; 1078 AA.
 AC Q9H2C6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Cellulose synthase.
 GN Name=Ces5;
 OS Populus tremuloides (Quaking aspen).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
 OX NCBI_TaxId=3693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Joshi C.P., Chavli R., Kalluri U.C., Samuga A.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RA Joshi C.P., Kalluri U.C.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 CC EMBL: AY055724; AA23710.2; -
 CC DR GO: GO:0016020; C:membrane; IEA.
 CC DR GO: GO:0016760; F:cellulose synthase (UDP-forming) activity; IEA.
 CC DR GO: GO:0030244; P:cellulose biosynthesis; IEA.
 CC DR InterPro: IPR005150; Cellulose synth.
 CC DR Pfam: PF03552; Cellulose synth; 1.
 CC SEQUENCE 1078 AA; 119539 MW; 3CDA85C184D7B53F CRC64;

Query Match 10.3%; Score 87.5; DB 2; Length 1078;
 Best Local Similarity 22.2%; Pred. No. 20;
 Matches 39; Conservative 23; Mismatches 65; Indels 49; Gaps 8;

19 LHETPAV-----LRLGPPGKAKWGMG---GMSFSLPQACVXGAAPMOTG-SPNRDGSQ 69
 178 VHGSNNRVVDVPAEFSGPGLGVAMKERVDMGKMDKTVV-----PMSTGHAPSERGAG 233

QY 70 GMDGA-----LGRGWTPLASCHPLROSALHPSPSLYESISGVCTTHPRCPSN 120
 DB 234 DIDATDVLDVDSILNDEARQPLSRKVSIPESKINPRMIVLGLICFLH-YRITNP 292

QY 121 TQHH-----WLRPNSLSQIRFTISIFINGETSPSH 152
 DB 293 VRNVALGLVIGWDMFAISRIIDQPPKMLPGNRETYLDRLALRYMBGE--PSH 346

RESULT 14
 06NZ15 PRELIMINARY; PRT; 1449 AA.

AC 06NZ15
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Collagen, type I, alpha 3.
 GN Name=coll1a3;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.U., Abramson R.D., Mullany S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fabley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzyzanski M.I., Skalska U., Smaltz D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC066384; AAH66384.1; -;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR000885; F1b_collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 17.
 DR Pfam; PF00093; VWC; 1.
 DR ProDom; PD000007; C1g_helix; 4.
 DR ProDom; PD002078; F1b_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01208; VWC_1; 1.
 DR PROSITE; PS0184; VWC_2; 1.
 KW Collagen.
 SQ SEQUENCE 1449 AA; 136980 MW; A763D6C12BDFD07 CRC64;

Query Match 10.2%; Score 86.5; DB 2; Length 1449;
 Best Local Similarity 22.6%; Pred. No. 34;
 Matches 44; Conservative 12; Mismatches 54; Indels 85; Gaps 9;

QY 30 GPPG-KAKGKGMWSFLPQACVXGAAPM-----QTGSPNRGSGGMDGALGPRG----- 79
 DB 1108 GPPGTSGEQG-----PAGATPAGRGFAGSNGSGKGNNGVGPVGPFGRGAS 1158

QY 80 -----WTPASCHPLROSALHPSP-----SLIYE 103
 DB 1159 GEMGPAGAPGLPGPPPPGASGPFPLPMPQKXGPDPLRGYRADASVDRDSEVDT 1218

QY 104 SIGSCVCTTHPRCPENTQ-----HYWLRN--SELSQIRFTSI 142
 DB 1219 SLKSLSQKXENTIRSPGTOANPARMCRDLRMCHPWSKSGSVWVDNQGSLDAIK-----V 1274

QY 143 FINGER-----SPSH 152
 DB 1275 FCNMTGFTCVSDPYH 1289

RESULT 15
 06PE19 PRELIMINARY; PRT; 1449 AA.

AC 06PE19
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Collagen, type I, alpha 3.
 GN Name=coll1a3;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.U., Abramson R.D., Mullany S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fabley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzyzanski M.I., Skalska U., Smaltz D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC058045; AAH58045.1; -;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002181; Fibrinogen_C.

DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF01410; COL1, 1.
 DR Pfam: PF01391; Collagen; 17.
 DR Pfam: PF00093; VWC; 1.
 DR ProDom: PD000007; C1g_helix; 4.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COL1; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF_C_1; 1.
 DR PROSITE: PS50184; VWF_C_2; 1.
 DR Collagen.
 SW SEQUENCE 1449 AA; 136930 MW; 9A51BA0B674B08B8 CRC64;

Query Match 10.2%; Score 86.5; DB 2; Length 1449;
 Best Local Similarity 22.6%; Pred. No. 34;
 Matches 44; Conservative 12; Mismatches 54; Indels 85; Gaps 9;

QY 30 GPRG-KAKMGWGWMSFSLPQACVXGAAPN---QTGSPNRDGSQGMGALGPRG----- 79
 Db 1108 GPRGTSGEQG-----PAGATGPAGPRGPAAGNSGSPKCDGMNGVPGVGPGRGRS 1158
 QY 80 -----WTPASCHLPLRQSLHPS-----SLAYE 103
 Db 1159 GEMGPAGAPGLPGPPGPGSAGPGEPPLMPQOEKGPDLRGYRADNASVRDRSEVDT 1218
 QY 104 SIGSVCTTHPYRCPSNTQ-----HYHMLRPN--SELSQIRFTISI 142
 Db 1219 SLKSLSGKIENTIRSPGCTQANPARMCRLMCHPEWKSQSYWVDPNQGSPLDAIK-----V 1274
 QY 143 FINGET-----SPSH 152
 Db 1275 FCNNETGETCVSPYH 1289

Search completed: October 24, 2004, 16:03:24
 Job time : 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:03:58 ; Search time 2704 Seconds

(without alignments)
2048.387 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1. MPHLYIDGVFPPIQVREKLN.....LSQIRFTISIFINGETSPSH 152

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09977418/runat_22102004_115948_1031/app_query.fasta_1.327
-DB=BS1 -QPMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowm62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -WINLEN=0 -MAXLEN=200000000
-USERS=US09977418 @CGN 1.1_3437 @runat_22102004_115948_1031 -MCPD=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

EST:
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	817	96.3	901	7	CF552903 AGNCOURT
2	777	91.6	1139	4	BM926609 AGNCOURT
3	765	90.2	974	2	BR250170 600924555
4	753	88.8	508	6	CB267862
5	752	88.7	878	5	EX402471 EX402471
6	748	88.2	726	2	BF792513 602253622
7	731	86.2	624	2	W61026
8	709	83.6	636	2	BF306561 601889886
9	683	80.5	496	1	AJ707312 AJ707312

10	666	78.5	376	6	CB267751 1006657 H
11	662	78.1	528	1	AJ711705
12	662	78.1	956	2	BF203383 601865995
13	662	78.1	976	2	BF203719 601866380
14	631	74.4	518	6	CD625679 560505765
15	633	68.8	562	4	BI113724 602860919
16	531	62.6	848	7	CF552003
17	527	62.1	605	2	AM384208
18	522	61.6	416	1	AJ711115
19	506	59.7	951	2	BF306631 601888886
20	494	58.3	501	1	AJ711059
21	487	57.4	443	2	BE840664
22	479.5	56.5	503	7	N55874
23	468	55.2	508	1	AA211656
24	462	54.5	409	2	BE840591
25	456	53.8	400	6	CD625675
26	456	53.8	400	6	CD625677
27	449	52.9	404	6	CD625674
28	439	52.9	409	6	CD625676
29	436	51.4	502	1	AJ708694
30	428.5	50.5	400	6	CD625678
31	418	49.3	389	6	CD4005
32	396	46.7	227	1	AA042804
33	372	43.9	395	6	CB050516
34	372	43.9	395	6	CB050517
35	372	43.9	395	6	CB050518
36	330	38.9	256	6	CO2777
37	330	38.9	690	5	BQ006459
38	322	38.0	681	5	BM936902
39	321	37.9	329	1	AI805232
40	321	37.9	329	1	AI817452
41	321	37.9	329	1	AI817826
42	318	37.5	333	1	AJ709557
43	306	36.1	183	1	AJ709804
44	301	35.5	236	7	F00226
45	288	35.1	672	6	CA424992

ALIGNMENTS

RESULT 1
CF552903 901 bp mRNA linear EST 22-SEP-2003
LOCUS AGNCOURT 15596050 NIH MGC 183 Homo sapiens cDNA clone
DEFINITION IMAGE:10529648 5', MENA sequence.

ACCESSION CF552903
VERSION CF552903.1 GI:34889737
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (base 1 to 901)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cga@nci.nih.gov

Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNI at:
http://image.llnl.gov
Plate: NDAM617 row: 1 column: 17
High quality sequence stop: 674.
Location/Qualifiers
1..901
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30529648"
/lab_host="DH10B-nona (T1 and T5 phage resistant)"
/clone_lib="NIH_MGC_183"
/note="Organ: Pooled muscle (cardiac and skeletal);
Vector: PCMV-Sports.1; Site_1: EcoRV (destroyed); Site_2:
NotI; Library is oligo-dt primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.7. Library was constructed by Invitrogen."

ORIGIN

Alignment Scores:

Pred. No.:	1,766-57	Length:	901
Score:	817.00	Matches:	150
Percent Similarity:	98.05%	Conservative:	1
Best Local Similarity:	97.40%	Mismatches:	1
Query Match:	96.34%	Indels:	2
DB:	7	Gaps:	0

US-09-977-418-8 (1-152) x CF552903 (1-901)

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 231 ATGCCACATCTGTATATAGATGGGTTTTCATATACAGCTGTTCTGATTAACCTGCAT 290
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTyrGly 40
DB 291 GAAACCTCCGCTCCCTGCGCTGCGGCTCCAGGCAAGCCAAAGTGGGTTGGGG 350
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 351 TGGGGCTGTCTCTCTCCCTCCACAGGCTGTGTTCTTGGGGCTGCTCCATGACACA 410
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 411 GGATCACCCTAACAGAGATGGAGCCAGGCGCATGATGAGGCTTGGGCTCTCGAGGTTGG 470
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 471 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGGAGTGTAGCTTCCATCCATCCCTCTTT 530
QY 100 uileTyGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
DB 531 AATCTATGATCATATAGGCTCGGTGTGTGAACACACACACCCATAGCTGTCTCTCA 590
QY 120 AsnThrGlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGlnLeuArgPheTrr 140
DB 591 AATCTACAGCATTAACCATGGTGTAGGCGCAATTCAGAGCTTCTCAATCAGATTACAA 650
QY 140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152
DB 651 TCTCCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 688

RESULT 2
BM926609 1139 bp mRNA linear EST 12-MAR-2002
LOCUS BM926609
DEFINITION AGENCOURT 6644683 NIH_MGC_122 Homo sapiens CDNA clone IMAGE:5767017
5', mRNA sequence.
ACCESSION BM926609
VERSION BM926609.1 GI:19376988
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1139)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNL412824 row: j column: 10
High quality sequence stop: 717.
Location/Qualifiers
1. 1139
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5767017"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: PCMV-Sports6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleen. Library is oligo-dt
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

FEATURES
source

ORIGIN

Alignment Scores:

Pred. No.:	4,866-54	Length:	1139
Score:	777.00	Matches:	148
Percent Similarity:	95.51%	Conservative:	1
Best Local Similarity:	94.87%	Mismatches:	3
Query Match:	91.63%	Indels:	4
DB:	4	Gaps:	0

US-09-977-418-8 (1-152) x BM926609 (1-1139)

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 398 ATGCCACATCTGTATATAGATGGGTTTTCATATACAGCTGTTCTGATTAACCTGCAT 457
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTyrGly 40
DB 458 GAAACCTCCGCTCCCTGCGCTGCGGCTCCAGGCAAGCCAAAGTGGGTTGGGG 517
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 518 TGGGGCTGTCTCTCTCCCTCCACAGGCTGTGTTCTTGGGGCTGCTCCATGACACA 577
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 578 GGATCACCCTAACAGAGATGGAGCCAGGCGCATGATGAGGCTTGGGTTCTCGAGGTTGG 637
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 638 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGGAGTGTAGCTTCCATCCATCCCTCTTT 697
QY 100 uileTyGluSerIleGlySerVal-CysValThr-ThrHisProTyrArgCysProSer 119
DB 698 AATCTATGATCATATAGGCTCGGTGTGTGAACACACACACCCATAGCTGTCTCTCA 757
QY 120 AsnThrGlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGlnLeuArgPheTrr 139
DB 758 AATCTACAGCATTAACCATGGTGTAGGCGCAATTCAGAGCTTCTCAATCAGATTAC 817
QY 139 IleSerIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 818 AATCTCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 857

RESULT 3
BE250170 974 bp mRNA linear EST 13-JUL-2000
LOCUS BE250170

```

DEFINITION 60094355F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960076 5',
mRNA sequence.
ACCESSION BE250170
VERSION BE250170.1 GI:9120277
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 974)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straube, Ph.D.
Email: ggaube-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LUCM52 row: n column: 13
High quality sequence stop: 692.
Location/Qualifiers
1..974
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2960076"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 17"
/note="Organ: muscle; Vector: pORF7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected by
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

ORIGIN

Alignment Scores:

Pred. No.:	3,89e-53	Length:	974
Score:	765.00	Matches:	147
Percent Similarity:	95.48%	Conservative:	1
Best Local Similarity:	94.84%	Mismatches:	4
Query Match:	90.21%	Indels:	4
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x BE250170 (1-974)

QY 1 MePProHsLeuYrTtLeaSGlYValPheProIleGlnLeuValArgGlnLeuHis 20

DB 211 ATGCCAATCTGATATAGATGAGGATTTTTCATACAGCTGGTCTGATACAGTGCAT 270

QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGly 40

DB 271 GAAATCTCTGCGCTCGCGCTCTGAGGCTCTCAAGCAAGGCAAGTGGGGTGGGG 330

QY 41 ThrGlyTrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 331 TGGGGGCTGCTCTCTCCCTCCACAGGCTGtTCTTGGGGCTGCTCCACAGCGACA 390

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80

DB 391 GGATCACTACACAGATGGAGCCA-GGCAATGATGGGGCTTGGGGCTCTCGAGGTGG 449

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100

DB 450 ACCCAAGCTCTTCCCAAGCTTCCCTCCGAGCAATGAGCTCTCCATCCATCCCTCTTT 509

QY 100 uiletGylusertleGlyserValCysValThr-ThrHisProGlyArgCysProsera 120

```

DB 510 AATCTATGATCTATAGGCTCGTGTGTATACACACACCCATGCTGTCTTCA 569
QY 120 snThrGlnHisTyrHisTrpLeuArgProAsnSer-GluLeuSerGlnIleArgPheThr 139
DB 570 ATACTCAGATTAACCATTTGGTTGAGGCCAATTCAGAGCTTTCTCAATCAGATTACA 629
QY 140 TleSerTlePheIleAsnGlyCiuThrSerProSerHis 152
DB 630 ATCTCATTTTCATTACAGGGGAAAAACATCCGAGCCAC 668
RESULT 4
CB267862
LOCUS CB267862 508 bp mRNA linear EST 20-FEB-2003
DEFINITION 1006768 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
ACCESSION CB267862
VERSION CB267862.1 GI:28442447
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 508)
AUTHORS Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
TITLE EST analysis of human adipose gene expression
JOURNAL Unpublished (2002)
COMMENT Contact: Gong Da-Mei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St, HH497, Baltimore, MD 21201, USA
Tel: 410 706 1672
Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu
PCR Primers
FORWARD: CTCGGGAAGCGGCCATTGTGTGCT
BACKWARD: AATACGACTCACTATAGAGGGAATTGG
Seq primer: GTTGGTACCGGGAATTC.
Location/Qualifiers
1..508
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="Male and Female"
/tissue_type="Adipose"
/clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
/note="Vector: lambdaTriplex"
```

ORIGIN

Alignment Scores:

Pred. No.:	1.63e-52	Length:	508
Score:	753.00	Matches:	139
Percent Similarity:	97.90%	Conservative:	1
Best Local Similarity:	97.20%	Mismatches:	1
Query Match:	86.80%	Indels:	2
DB:	6	Gaps:	0

US-09-977-418-8 (1-152) x CB267862 (1-508)

QY 12 TleGlnLeuValArgGlnLysLeuHisGlnThrProAlaValLeuArgLeuGlyPro 31

DB 1 ATACAGCTGCTCGATATACATGATTAACCTCTCCGCTCTCGCTGAGGGCT 60

QY 32 ProGlyLysAlaLysTTPGlyTTPGlyTTPGlyTrpSerPheSerLeuProGlnAlaCys 51

DB 61 CAGGCAAGGCAATGGGGTGGGGTGGGGCTGCTCTCTCCCTCCACAGGCTGT 120

QY 52 Val***GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnLysT 71

DB 121 GTTCTGGGGCTGCTCCCAATGCAAGATCACTTAACAGATGGAGCCAGGCAATG 180

QY 72 AspGlyAlaLeuGlyProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg-Gl 91

DB 181 GATGGGGCTTGGTCTCGAGGTGAGACCCAGCTTCTTGGCACCTTCCCTCCGGACA 240

QY 91 nserAlaLeuHisProSerProSerLeuLeuIleTyrGluSerIleGlySerValCysValTh 111

DB 241 GTGAGCTCTCCACCAACCCCTTTATCTATGATCTATAGAGCTGGGTGTGTAAC 300

QY 111 T-ThrHisProTyrArgCysProSerAsnThrGlnHisTyrHisIleTyrLeuArgProAsn 131

DB 301 AACACACACCCCTATCGTTTCCCTTCAATACACAGATTACCATTTGGTTGAGCCAAATT 360

QY 131 eTGLuLeuSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGluThrSerProS 151

DB 361 CAGAGCTTCTCATAAATCAGATTATTCATCTTCATTAATACGGGGAACATCCCGCA 420

QY 151 exHis 152

DB 421 GCCAC 425

RESULT 5
BX402471 878 bp mRNA linear EST 29-APR-2004
LOCUS BX402471 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
DEFINITION clone CS0D1039YB13.3-PRIME, mRNA sequence.
ACCESSION BX402471
VERSION BX402471.2 GI:46874966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 878)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30618825.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 131 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr; Web: www.genoscope.cns.fr
Description: This sequence was derived from a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
545.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdnas=CS1A1010ZF07NP1&c=545.r.
FEATURES
source
1..878
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039YB13"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1lb="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 4.04e-52 Length: 878
Score: 752.00 Matches: 135
Percent Similarity: 94.44% Conservative: 1
Best Local Similarity: 93.75% Mismatches: 0
Query Match: 88.68% Indels: 0
Gaps: 0
US-09-977-418-8 (1-152) x BX402471 (1-878)

QY 9 ValPheProIleGluLeuValArgGluLysLeuHisGluThrProAlaValLeuArgPhe 28
|||||
:::
|||||

DB 876 GTTTTCCAAATACAGCTGCTTCGTGATTAACATGACATGAACCTCCGCTCCGCCCTG 817

QY 29 LeuGlyProProGlyValAlaValTyrPheGlyTyrGlyTyrPheSerPheSerLeuPro 48

DB 816 CTGGAGCTCTCAGACAGACAGCAAGCTGGAGGTGGGGCTGGCTCTTCTCCCTCCA 757

QY 49 GlnAlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySer 68

DB 756 CAGGCTGTGTCTTGGGGCTGCTCCATCAGACAGATTAACCTTAACAGATGGAAC 697

QY 69 GlnGlyMetAspGlyAlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuPro 88

DB 696 CAGGCGATGGATGGGCTTGGGTCCTCAGAGTGAACCCAGCTCTTGGCACCTTCCC 637

QY 89 LeuArgInsSerAlaLeuHisProSerProSerLeuIleTyrGluSerIleGlySerVal 108

DB 636 TCCGGGAGTCACCTCCATTCATCCATCCCTCTTATCATATCATATAGCTCGTGT 577

QY 109 CysValThrThrHisProTyrArgCysProSerAsnThrGlnHisTyrHisIlePhe 128

DB 576 GTGTMACACACACACCCCTATCGTTTCCCTTCAAAATCAGATTACCATTTGGTTGAG 517

QY 129 ProAsnSerGluLeuSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGluThr 148

DB 516 CCAATTCCAGCTTCTCATAAATCAGATTATTCATCTTCATTTCAATACGGGGAACA 457

QY 149 SerProSerHis 152

DB 456 TCCCCGAGCCAC 445

RESULT 6
BF792513 726 bp mRNA linear EST 12-JUN-2001
LOCUS BF792513
DEFINITION 602253622F1 NIH_MGC_84 Homo sapiens CDNA clone IMAGE:4345876 5',
mRNA sequence.
ACCESSION BF792513
VERSION BF792513.1 GI:12097567
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 726)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL Unpublished (1999)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: L1AM966 row: 1 column: 05
High quality sequence step: 724.
FEATURES
source
1..726
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4345876"
/issue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 84"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	6,73e-52	Length:	726
Score:	748.00	Matches:	146
Percent Similarity:	94.84%	Conservative:	1
Best Local Similarity:	94.19%	Mismatches:	5
Query Match:	88.21%	Indels:	4
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x BF792513 (1-726)

QY 1 MetProHisLeuYrYrIleAspGlyValPheProIleGlnLeuValArgGlyLysLeuHis 20
 Db 151 ATGCCACATCTGTATATAGATGGGCTTTTTCATACAGCTGGTCTGTATTAACATGCAT 210

QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGlyTTPGly 40
 Db 211 GAAATCTCTGCGCGCTCCGCTCTGCGCTCTCCAGCAAGCCACGCTGGGCTTGGGTG 270

QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 271 TGGAGCTGGTCTCTCTCTCCCTCCACAGGCTGTCTTCTGGGCTGTCTCCATGCAGACA 330

QY 61 GlySerProAlaArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 331 GGATCACCTAAACAGATGAGCAAGCA-GGCATGATGGGCTTGGGCTCTCCAGGTTGG 389

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 390 ACCCAGCTCTTCCACCTCTCCCTCCGAGCTCAGCTCTCCATCCATCCCTCTTT 449

QY 100 uileTyGluSerIleGlySerValCysValThr-Thr-HisProTyArgCysProSer 119
 Db 450 AATATGAAATCTATAGGCTCGCTGTGTATACACACACCCCTATCTGTCTCTTCA 509

QY 120 AsnThrGlnHisTyrHisTTPLeuArgProAsnSerGluLeuSerGlnIleArgPheThr 139
 Db 510 AATCTCAGCATTCATCCATGGTGTAGGCCAATTCACAGCTTCTCTCAATCAAGATTACA 569

QY 140 IleSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 570 ATCTCCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 608

RESULT 7

W61026 624 bp mRNA linear EST 11-OCT-1996
 LOCUS zc54c09.1 Soares senescent fibroblasts_NBHSF Homo sapiens cDNA
 DEFINITION clone IMAGE:326128 5', mRNA sequence.

W61026 61367823
 VERSION W61026.1 GI:1367823

KEYWORDS EST.
 ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 624)
 Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.

TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNC; contact the
 IMAGE Consortium (info@image.lln.gov) for further information.
 Insert Length: 962 Std Error: 0.00
 Seq primer: ETPRimer
 High quality sequence stop: 447.

FEATURES

source

Location/Qualifiers

1..624
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="Gene:1257641"
 /db_xref="taxon:9606"
 /clone="IMAGE:326128"
 /tissue_type="senescent fibroblast"
 /lab_host="DH108 (ampicillin resistant)"
 /clone_lib="Soares senescent fibroblasts_NBHSF"
 /note="Vector: pT7T3D (Pharmacia) with a modified
 polylinker V type; phagemid; site 1: Not I; site 2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer 15',
 TTTTACCATCTGAAGTGGAGCGGCGCATTTTCTTTTCTTTTCTTTT 3',
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Fatima Bernaldo."

ORIGIN

Alignment Scores:

Pred. No.:	1.41e-50	Length:	624
Score:	731.00 <td>Matches:</td> <td>134</td>	Matches:	134
Percent Similarity:	97.83% <td>Conservative:</td> <td>1</td>	Conservative:	1
Best Local Similarity:	97.10% <td>Mismatches:</td> <td>2</td>	Mismatches:	2
Query Match:	86.20% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	7	Gaps:	0

US-09-977-418-8 (1-152) x W61026 (1-624)

QY 17 GlnYrLeuHisGlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLys 36
 Db 1 GATTAACATGCAATGAATCTCTGCGCTCCGCTCGGCGCTCCACAGCAAGCCAAAG 60

QY 37 TTPGlyTTPGlyTTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAla 56
 Db 61 TGGGCTGGGGGTGGGGGCTGTCTCTCCCTCCACAGGCTGTCTTGGGCTGT 120

QY 57 ProMetGlnThrGlySerProAlaArgAspGlySerGlnGlyMetAspGlyAlaLeuGly 76
 Db 121 CCCATGCAACAGAGATCACTTAACAGATGAGAACCCAGGATGATGGGCTTGGGT 180

QY 77 ProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisPr 96
 Db 181 CCTCAGGTTGACCCCGAGCTTCTTCCACCTTCCCTCCGAGCTCAGCTCTCATTCC 240

QY 96 oSerProSerLeuIleTyGluSerIleGlySerValCysValThr-ThrHisProTyArg 116
 Db 241 ATCCCCCTTTAATCTATGATCTATAGGCTCGGTGTGTATACACACACCCCTATC 300

QY 116 rGySerProSerAsnThrGlnHisTyrHisTTPLeuArgProAsnSerGluLeuSerGlnI 136
 Db 301 GTTGCTCTCAATACTCGCATTCATTCATGATGTAGGCAATTCACAGCTTCTCATA 360

QY 136 lAsrPheThrIleSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 361 TCAGATTTCATATTCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 410

RESULT 8

BF306561 636 bp mRNA linear EST 21-NOV-2000
 LOCUS 601888986F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122788 5',
 DEFINITION mRNA sequence.

W61026 61367823
 VERSION BF306561.1 GI:11253663

KEYWORDS EST.
 ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 636)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: image.llnl.gov
Plate: LLCM1006 row: 1 column: 21
High quality sequence stop: 634.
Location/Qualifiers

FEATURES
source

1..636
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4122788"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 17"
/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 9,51e-49 Length: 636
Score: 709.00 Matches: 131
Percent Similarity: 97.06% Conservative: 1
Best Local Similarity: 96.32% Mismatches: 3
Query Match: 83.61% Indels: 2
DB: Gaps: 0

US-09-977-418-8 (1-152) x BF306561 (1-636)

QY 1 MetProHisLeuTyrTleAspGlyValPheProIleGlnLeuValArgGluLeuHis 20
DB 230 ATGCCACATCTGATATAGATGGGTTTTCATACACCTGGTTCGATTAACCTGCAT 289
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
DB 290 GAAACTCTGCGCTCTGCGCTGCGGCTCCAGGCAAGCCAAAGGGGTTGGGG 349
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
DB 350 TGGGGCTGCTCTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCATGACACA 409
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaGlyProArgGlyTyr 80
DB 410 GGATCACCTTAACAGATGGAAGCCA-GGCATGATGGGGCTTGGGCTCTGAGCTGG 468
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 469 ACCCGAGCTTCTTGACACCTCCCTCCGCGACAGCTCCATCCATCCCTCTTT 528
QY 100 uTleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAs 120
DB 529 AATCTATGATCTATAGCTCGCTGCTGTATACACACACCCCTATCTGTCTTCAA 588
QY 120 nThrGlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGln 135
DB 589 TACTCAGCATTAACATGGTGTAGGCGCAATTCAGAGCTTCTCTCAA 634

RESULT 9
AJ707312/C

LOCUS AJ707312 496 bp mRNA linear EST 30-JUN-2004
DEFINITION AJ707312 CPM01 Homo sapiens cdna clone CPM01259, mRNA sequence.
ACCESSION AJ707312
VERSION AJ707312.1 GI:49501607
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 496)

AUTHORS Mallino C.
TITLE Expression profiling of human hypertrophic cardiomyopathy progressed to dilated cardiomyopathy by DNA microarray analysis

JOURNAL
COMMENT

Contact: Mallino C
Biology and CRI
University of Padova
Via U. Bassi, 58/B, 35131, ITALY.
Location/Qualifiers

FEATURES
source

1..496
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CPM01259"
/tissue_type="heart"
/clone_lib="CPM01"
/note="caucasian"

ORIGIN

Alignment Scores:

Pred. No.: 9,69e-47 Length: 496
Score: 683.00 Matches: 128
Percent Similarity: 96.24% Conservative: 0
Best Local Similarity: 96.24% Mismatches: 2
Query Match: 80.54% Indels: 3
DB: Gaps: 0

US-09-977-418-8 (1-152) x AJ707312 (1-496)

QY 23 ProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTrrGly 42
DB 486 CTGCGCTCTGCGCTCTGCGGCTCCAGGCAAGCCACCGTGGGTTGGGGCTGGGC 427
QY 43 TrrSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThrGlySer 62
DB 426 TGGTCTCTCTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCATGACAGATCA 367
QY 63 ProAsnArgAspGlySerGlnGlyMetAspGlyValAlaGlyProArgGlyTrrThrPro 82
DB 366 CTTAACAGAGATGGAAGCCAGGCGCATGATGGGGCTTGGGCTCTGAGGTTGACCCCA 307
QY 83 AlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLeuTyr 102
DB 306 GCTTCTTGCCACCTTCCCTCCGCGAGTACGCTCTCCATCCATCCCTCTTTATCTA 247
QY 102 rGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerAsnThr 121
DB 246 TGAATCATATAGCTCGGTGTATACACACACACCCATGCTGTCTTCAATATCTA 187
QY 122 GlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrIleSer 141
DB 186 CACCATTAACCATGTTGAGGCCAATTCAGAGCTTCTTCAATCAATTCATTC 127
QY 142 IlePheIleAsnGlyGlnThrSerProSerHis 152
DB 126 ATTTTCACTAACCGGGAACATCCCGAGCCAC 94

RESULT 10
CB267751 376 bp mRNA linear EST 20-FEB-2003
LOCUS CB267751
DEFINITION 100657 Human Fat Cell 5'-Stretch Plus cdna Library Homo sapiens
CDNA 5', mRNA sequence.
ACCESSION CB267751

VERSION CB267751.1 GI:28442336
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 376)
 Yang, R.-Z., Shuidiner, A. and Gong, D.-W.
 EST analysis of human adipose gene expression
 Unpublished (2002)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Division of Endocrinology, Diabetes and Nutrition
 University of Maryland
 660 Redwood St, H497, Baltimore, MD 21201, USA
 Tel: 410 706 1672
 Fax: 410 706 1622
 Email: dgong@medicine.umaryland.edu

PCR PRIMERS
 FORWARD: CTCGGGAAGCGCCCATGTGTGCT
 BACKWARD: AATGAGTCACTAATGAGCGAATTGG
 Seq primer: GTTGTACCCGGAATTC.

FEATURES
 source
 1..376
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /sex="Male and Female"
 /tissue_type="Adipose"
 /clone_idb="Human Fat Cell 5'-Stretch Plus cDNA library"
 /note="Vector: lambdaTriplex"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,71e-45 Length: 376
 Score: 666.00 Matches: 122
 Percent Similarity: 97.62% Conservative: 1
 Best Local Similarity: 96.83% Mismatches: 2
 Query Match: 78.54% Indels: 1
 DB: Gaps: 0

US-09-977-418-8 (1-152) x CB267751 (1-376)

QY 13 GlnleuValArgGlnLeuLeuGlnThrProAlaValLeuArgLeuGlnProPro 32
 Db 3 CAGCTGGTGGTGAATAAGTCATGAAGTCTCTCCGCTCGGCGCTGCGGCGCTCCA 62

QY 33 GlyLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGlnAlaCysVal 52
 Db 63 GGCAGAGCCAGAGTGGGGTGGGGCTGGGCTGCTCCCTCCACAGGCGCTGTGT 122

QY 53 **GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyMetAsp 72
 Db 123 CTGGGGGCTGCTCCATGACAGACAGATCACTAACAGAGATGGAAGCCAGGGCATGAT 182

QY 73 GlyAlaLeuGlyProArgGlyTrpThrProAlaSerCysHisLeuProLeuArg-Glnse 92
 Db 183 GGGGCTTGGGCTCTCGAGGTGGAGCCCACTTCTTGCAACCTTCCCTCCGCGGCGAGTC 242

QY 92 AlaLeuHisProSerProSerLeuLeuIleTyrgLuserIleGlySerValCysValThr-T 112
 Db 243 AGCTTCCATCCATCCCCCTTTTAATCTAATAGCTAATAGCTCGGTGGTGTAAACA 302

QY 112 hrHisProTyArgCysProSerAsnThrGlnHisIleTyrgLuserIleArgProAsnSerg 132
 Db 303 CACACCCCTATCGTTGCTTCAATATCAATGATCACTTCAATGTTGAGGCGCAATTCAG 362

QY 132 LuLeuSergIle 136
 Db 363 AGCTTTTCAATC 376

RESULT 11
 AJ711705 528 bp mRNA linear EST 30-JUN-2004
 LOCUS

DEFINITION AJ711705 CMPD01 Homo sapiens cDNA clone CMPD09581, mRNA sequence.
 ACCESSION AJ711705
 VERSION AJ711705.1 GI:49506000
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 528)
 Millino, C.
 Expression profiling of human hypertrophic cardiomyopathy
 progressed to dilated cardiomyopathy by DNA microarray analysis
 Unpublished (2004)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Millino C
 Biology and CRIBI
 University of Padova
 Via U. Bassi, 58/B, 35131, ITALY.

FEATURES
 source
 1..528
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CMPD09581"
 /tissue_type="heart"
 /clone_idb="CMPD01"
 /note="Caucasian"

ORIGIN
 Alignment Scores:
 Pred. No.: 5.74e-45 Length: 528
 Score: 662.00 Matches: 121
 Percent Similarity: 96.83% Conservative: 1
 Best Local Similarity: 96.03% Mismatches: 2
 Query Match: 78.07% Indels: 2
 DB: Gaps: 0

US-09-977-418-8 (1-152) x AJ711705 (1-528)

QY 29 LeuGlyProProGlyLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuPro 48
 Db 22 CTGGGGCTCTCCAGGAGCCAGAGTGGGGTGGGGTGGGGTGGGCTCTCCCTCCCA 81

QY 49 GlnAlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySer 68
 Db 82 CAGGCTGTGTCTTGGGCTGCTCCATGACAGATCACTCAACAGATGGAAGC 141

QY 69 GlnGlyMetAspGlyAlaLeuGlyProArgGlyTrpThrProAlaSerCysHisLeuPro 88
 Db 142 CAGGCGATGATGGGGCTTGGGCTCTCGAGGTGGAGCCCACTTCTTGCCACCTTCCC 201

QY 89 LeuArg-GlnSerAlaLeuHisProSerProSerLeuLeuIleTyrgLuserIleGlySerVal 108
 Db 202 CTCCGGGCGAGTCACTTCCATCCATCCCTCTTAACTAATCTAATGATCTAAGGCTCGGT 261

QY 108 LcyValThr-ThrHisProTyArgCysProSerAsnThrGlnHisIleTyrgLuserIle 128
 Db 262 GTGTGAACACACACACCCCTATCGTTGCTTCAAAATCAAGCATTAACCATGTGTA 321

QY 128 rgProAsnSergLuLeuSergGlnIleTyrgPheThrIleSertIlePheLeuSergIle 148
 Db 322 GGCCAATTCAGAGCTTCTCAATATCAATATCAATATTCATTTCAATTAAGGGGANA 381

QY 148 hrSerProSerHis 152
 Db 382 CATCCCGAGGCAC 395

RESULT 12
 BF203383 956 bp mRNA linear EST 06-NOV-2000
 LOCUS BF203383
 DEFINITION 601865955F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4096646 5',
 mRNA sequence.
 ACCESSION BF203383
 VERSION BF203383.1 GI:11096969

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 956)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at: image.llnl.gov
Plate: LLCM965 row: n column: 23
High quality sequence stop: 682.

FEATURES

source
1..956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:409846"
/issue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7, Site_1: EcoRI;
Site_2: XhoI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 1,256-44 Length: 956
Score: 662.00 Matches: 131
Percent Similarity: 91.78% Conservative: 12
Best Local Similarity: 89.73% Mismatches: 3
Query Match: 78.07% Indels: 0
Gaps: 2

US-09-977-418-8 (1-152) x BF203383 (1-956)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 225 ATGCCACATCTGATATAGATGGGGTTTTCATAACAGCTGCTGATTAACATGCAT 284
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGly 40
DB 285 GAAACTCTGCGCTCTGCGCTGCGCTGCGGCTCCAGGAGGCCAGCGGGGTTGGGG 344
QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 345 TGGGGCTGGCTCTTCCCTCCACAGGCTGTGTTGGGGCTGCTCCATGCAGACA 404
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetLaspGlyAlaLeuGlyProArgGlyTyr 80
DB 405 GATACCTTAACAGACATGAAAGCCA-GGCATGAGATGGGGCTTTGGTCTTGAGAGTTGG 463
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 464 ACCCGAGCTTCTTGCCACCTTCCCTCCGGGAGAGCAGCTCCATCCATCCCTCTTTA 523
QY 101 TleTyrGlnSerTlleGlySerValCysValThrThrHisProTyrArgCysProSerSer 120
DB 524 ATCTAAGATCTAAGAGCTGGTGTGTAAACAACACCC-TATCGTTGCTCTTCAAT 582
QY 121 ThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140

DB 583 ACTCAGATTAAC-CACTTGTTGAGGCCAATTCAGAGCTTCTCACAATCAATTAAATC 641
QY 141 SerIlePheIleAsnGly 146
DB 642 TCCATTATTAATTAACGGGA 659

RESULT 13
BF203719 976 bp mRNA linear EST 06-NOV-2000
LOCUS 601866380F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4099458 5',
DEFINITION mRNA sequence.
BF203719
ACCESSION BF203719.1 GI:11097305
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS 1 (bases 1 to 976)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at: image.llnl.gov
Plate: LLCM967 row: p column: 19
High quality sequence stop: 676.

FEATURES

source

1..976
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4099458"
/issue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7, Site_1: EcoRI;
Site_2: XhoI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 1,296-44 Length: 976
Score: 662.00 Matches: 137
Percent Similarity: 90.85% Conservative: 2
Best Local Similarity: 89.54% Mismatches: 13
Query Match: 78.07% Indels: 6
Gaps: 0

US-09-977-418-8 (1-152) x BF203719 (1-976)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 229 ATGCCACATCTGATATAGATGGGGTTTTCATAACAGCTGCTGATTAACATGCAT 288
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGly 40
DB 289 GAAACTCTGCGCTCTGCGCTGCGCTGCGGCTCCAGGAGGCCAGCGGGTGGGG 348
QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 349 TGGGGCTGGCTCTTCCCTCCACAGGCTGTGTTGGGGCTGCTCCATGCAGACA 408

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 409 GGATCACCTTAACAAGATGAGAGCA-GGCATGATGGGGC-TTGGGCTCTCGAGGTTGG 466
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 467 ACCCAAGCTTTTGGCACTTCCCTCCGGGAGATCAGCTCTCCATCCATCCCTCTTT 526
 QY 100 uUleTyGlnSerIleGlySerValCysValThr-ThiHisProTyArgCysProSerAs 120
 DB 527 AATCTATGAATCTATAGGCTCGGTGTGTGTATACACACACCC-TATCGTTGTGCTTCAA 585
 QY 120 nThGlnHisTyHisThrLeuArgProAsnSerGlnLeuSerGlnIleArgPheTrrI 140
 DB 586 TACTAGCATTAACATGATGGTGTAGGCGCAT-TCAAGACT-TCTCAAAATCAGATTTCAT 643
 QY 140 eSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 DB 644 CTGATTCTCATTAACGGGAGACAGATCCCGAGCCAT 680

RESULT 14

LOCUS CD625679 518 bp mRNA linear EST 12-JAN-2004
 DEFINITION 5605057601 FLP Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD625679
 VERSION CD625679.1 GI:40273945
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 TITLE 1 (bases 1 to 518)

COMMENT Circular rapid amplification of cDNA ends for high-throughput
 extension cloning of partial genes

JOURNAL Genomics 84 (1), 205-210 (2004)
 CONTACT: Fu GK
 INCYTE Genomics, Inc.
 3160 Porter Dr., Palo Alto, CA 94304, USA
 Tel: 6508454102
 Email: gfu@incyte.com

FEATURES

Location/Qualifiers

1..518
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="FLP"
 /note="Vector: pDrive Cloning Vector"

ALIGNMENT SCORES:

Pred. No.: 2,05e-42 Length: 518
 Score: 631.00 Matches: 117
 Percent Similarity: 95.93% Conservative: 1
 Best Local Similarity: 95.12% Mismatches: 3
 Query Match: 74.41% Indels: 2
 DB: 6 Gaps: 0

US-09-977-418-8 (1-152) x CD625679 (1-518)

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
 DB 145 ATGCCAATCTGATATAGATGGGGTTTTCATATACAGTGGTTCGTATAAAGTGCAT 204
 QY 21 GlnThrProAlaValLeuArgLeuGlyProProGlyLysAlaLysTrpGlyTrpGly 40
 DB 205 GAACCTCCTGCGCTGCGCTGCGCTGCGGCGCTCAAGCAAGCGACGTTGGGGG 264
 QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 DB 265 TGGGGCTGGTCTTCTCCCTCCACAGGCGCTGTCTTGGGGCTGCTCCCATGACAGACA 324

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 325 GGATCACCTTAACAAGATGAGAGCA-GGCATGATGGGGC-TTGGGCTCTCGAGGTTGG 384
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 385 ACCCAAGCTTTTGGCACTTCCCTCCGGGAGATCAGCTCTCCATCCATCCCTCTTT 444
 QY 100 uUleTyGlnSerIleGlySerValCysValThr-ThiHisProTyArgCysProSerAs 120
 DB 445 AATCTATGAATCTATAGGCTCGGTGTGTGTATACACACACCCATGCTAGTCTTCAA 504
 QY 120 snThr 121
 DB 505 ATRACT 509

RESULT 15

LOCUS B113724 562 bp mRNA linear EST 26-JUN-2001
 DEFINITION 602860919P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5020037 5',
 mRNA sequence.

ACCESSION B113724
 VERSION B113724.1 GI:14564625
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 TITLE 1 (bases 1 to 562)

COMMENT NIH-MGC http://mgi.mol.nh.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTRACT: Robert Strausberg, Ph.D.
 Email: ggebbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
 http://image.llnl.gov
 Plate: D1CML856 row: f column: 06
 High quality sequence stop: 562.

FEATURES

Location/Qualifiers

1..562
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /IMAGE="IMAGE:5020037"
 /clone_lib="NIH MGC 17"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOT7; Site 1: ECORI;
 Site 2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGGACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

ALIGNMENT SCORES:

Pred. No.: 2.13e-38 Length: 562
 Score: 583.00 Matches: 107
 Percent Similarity: 96.43% Conservative: 1
 Best Local Similarity: 95.54% Mismatches: 3
 Query Match: 68.75% Indels: 1
 DB: 4 Gaps: 0

US-09-977-418-8 (1-152) x B113724 (1-562)

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20

```

Db      226 ATGCCACATCTGTATATAGATGGGGTTTTCATACAGCTGCTCGTATAACTGCAT 285
Qy      21  GIUTHPICAlaValleuArgLeuGlyProProGlyLysAlaLysTyrGly 40
Db      286 GAAACTCCTGCCGCTCTGCGCTGCTGGGGCTCCAGGCCAGCCAGTGGGGTTGGGG 345
Qy      41  TPGlyTyrSerPheSerLeuPProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      346 TGGGGCTGGTCTCTCTCCCTCCACAGGCTGTCTTGGGGCTGCTCCCATGCAACA 405
Qy      61  GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db      406 GGATCACCTACAGAGATGAAAGCCAGGGCATGATGGGGCTTAGGGTCTCGAGGTTGG 465
Qy      81  ThrProAlaSerCysHisLeuPProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db      466 ACCCCAGCTTCTTGCCACCTTCCCTCCGGGCAAGTCAGCTCTCCATCCATCCCTCTTT 525
Qy      100 uIeTyrGlnSerIleGlySerValCysValThr 111
Db      526 AATCTATGAATCTATAGGCTCGGTGTGTATACA 559

```

Search completed: October 24, 2004, 17:57:19
 Job time : 2710 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:03:27 ; Search time 3547 Seconds
(without alignments)
2026.511 Million cell updates/sec

Title: US-09-977-418-8
Perfect score: 848
Sequence: 1 MPHLYIDGVFPIQLVREKJH.....LSQIRPTISIFINGETSPSH 152

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+;p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.spool/US09977418/runatc.22102004.115947.1017/app.query.fasta_1.327
-DB=GenBml -OPMT=fastap -SUFFIX=ge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=pl0sum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPRM=prc -NOR=exc -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09977418 -CGN_1_1_3731 -runatc.22102004.115947.1017 -NCPU=6 -ICPU=3
-NO_WAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

(GenBml):
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	846	99.8	1987	6	AX058600 Sequence
2	817	96.3	2152	9	AF155908 Homo sapi
3	817	96.3	2152	9	HS0243191 Homo sapi
4	817	96.3	2747	9	HS0803468 Homo sapi

C	5	817	96.3	109377	9	AL355994 Human DNA
C	6	811	95.6	2138	6	C0721116 Sequence
C	7	811	95.6	2192	6	AX017502 Sequence
C	8	811	95.6	2192	6	BD135192 Sequence
C	9	811	95.6	2589	6	BC006319 Homo sapi
C	10	811	95.6	2701	6	CQ783843 Sequence
C	11	811	95.6	2701	6	BD127861 Primer fo
C	12	811	95.6	2701	9	AK075151 Homo sapi
C	13	807	95.2	2142	9	AK095911 Homo sapi
C	14	807	95.2	76241	9	HS013364 Sequence
C	15	802	94.6	2410	6	AX747447 Sequence
C	16	802	94.6	2410	9	AK092275 Homo sapi
C	17	797	94.0	2163	9	BS337990 Homo sapi
C	18	772.5	91.1	601	11	EV167837 Sequence
C	19	772.5	91.1	2303	11	EV178365 Sequence
C	20	578	68.2	1550	6	AX081428 Sequence
C	21	512	60.4	305	6	AX058658 Sequence
C	22	382	42.7	201	11	EV203380 Sequence
C	23	233	27.5	306769	2	AL356105 Homo sapi
C	24	230	27.1	750	11	EV166340 HSPB7_990
C	25	112.5	13.3	188317	2	AC148957 Oclemtur
C	26	112.5	13.2	105628	2	AC138144 Homo sapi
C	27	111.5	13.1	62295	2	AC135730 Homo sapi
C	28	111.5	13.1	153152	9	AC021939 Homo sapi
C	29	111.5	13.1	161612	9	AC020754 Homo sapi
C	30	111.5	13.1	161629	2	AC087604 Pan trogl
C	31	111.5	13.1	209168	9	AC090955 Homo sapi
C	32	111	13.1	180876	9	AC021317 Homo sapi
C	33	110	13.0	123925	9	AC003976 Homo sapi
C	34	110	13.0	163628	2	AC139367 Homo sapi
C	35	110	13.0	177255	2	AC148806 Homo sapi
C	36	110	13.0	193646	9	AC131056 Homo sapi
C	37	109.5	12.9	208596	2	AC127589 Mus muscu
C	38	108.5	12.8	4268	9	HUMALPDP
C	39	108.5	12.8	143929	10	AC124714 Mus muscu
C	40	108.5	12.8	165196	10	AL596123 Mus muscu
C	41	108.5	12.8	169247	2	AL669970 Human DNA
C	42	108.5	12.8	343504	2	AL158031 Homo sapi
C	43	108	12.7	96678	9	AL096755 Homo sapi
C	44	108	12.7	129989	9	AP006621 Homo sapi
C	45	108	12.7	178778	9	AC132936 Homo sapi

ALIGNMENTS

RESULT 1	AX058600	1987 bp	DNA	linear	PAT 17-JAN-2001
LOCUS	AX058600	Sequence 7 from Patent WO0075321.			
DEFINITION	AX058600				
ACCESSION	AX058600.1	GI:12310942			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
	Homo sapiens (human)				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS	Shinkels,R.A., Fernandes,E., Herman,J. and Verner,C.				
TITLE	Polynucleotides and membrane-bound polypeptides encoded thereby				
JOURNAL	Patent: WO 0075321-A 7 14-DEC-2000;				
CURATOR	Cutagen Corporation (US)				
FEATURES					
source	location/Qualifiers				
	1..1987				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
	991..1449				
	/note="unamed protein product"				
	/codon_start=1				
	/protein_id="CAC22515.1"				
	/db_xref="gi:12310942"				
	/translation="MPHLYIDGVFPIQLVREKJHETPAVLELGGPKAKMGKMGMS				
	FSLPQACTVXGAHPVQGTSGPNRDSQMDGALGPRGWTIPASCHLPLRQSLAPSPSLIY				

CDS

ORIGIN

ESIGSVCTTHPYPSPNTQHYHWRPNSELSQIRFTTISIFINGETSPSH"

Alignment Scores:

Pred. No.:	7, 278-53	Length:	1987
Score:	846.00	Matches:	152
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.76%	Indels:	0
DB:	6	Gaps:	0

US-09-977-418-8 (1-152) x AX058600 (1-1987)

QY 1 MetProHisLeuTyrTlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20

DB 991 ATGCCACACCTGATATAGATGGGGTTTTCACATACAGCTGGTTGTAAGAAAACCTGAT 10550

QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGly 40

DB 1051 GAAATCCCTGCGCTCTGCGCTGCTGCGGCTCCACGCGAAGCCCAAGTGGGGTGGGG 11110

QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 1111 TGGGGCTGGCTCTTCCCTCCACAGGCTGTGTCNTGCGGGCTGCTCCATCGACAGACA 11770

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrrp 80

DB 1171 GGATCACCTTAACAGATGGAAGCCAGGCGATGATGGGGCTTTGGGTCCTCGAGGTGG 12330

QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100

DB 1231 ACCCCAGCTTCTTGCCACCTTCCCTCCGCGAGTCAGGCTCTCATCCATCCCTCTTTA 12990

QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerHis 120

DB 1291 ATCTAAGATCTATAGCTCGGTGGTGTGTAACAACAACCCCTATCGTGTCTTCAAT 13550

QY 121 ThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140

DB 1351 ACCACACATTAACATGCTGTGAGGCCAATTCCAGAGCTTTCATATCGATTTTCAATC 14110

QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152

DB 1411 TCCATTTTCATTAAACGGGAAACATCCCGACGACC 1446

RESULT 2 AFI55908 2152 bp mRNA linear PRI 27-DEC-1999

LOCUS Homo sapiens cardiovascular heart shock protein mRNA, complete cds.

ACCESSION AFI55908.1 GI:6635998

VERSION AFI55908.1 GI:6635998

KEYWORDS Homo sapiens (human)

SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 2152) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Krieff,S., Faivre,J.F., Robert,P., Le Douarin,B.,

TITLE Brunet-Latignon,N., Lefevre,I., Bouzyk,M.M., Anderson,K.M.,

Grellier,L.D., Tobin,F.L., Souchet,M. and Brill,A.

JOURNAL Identification and characterization of cHSP. A novel human small

REFERENCE stress protein selectively expressed in cardiovascular and

MEDLINE insulin-sensitive tissues

PUBMED J. Biol. Chem. 274 (51), 36592-36600 (1999)

AUTHORS Krieff,S., Faivre,J.F., Robert,P., Le Douarin,B.,

TITLE Brunet-Latignon,N., Lefevre,I., Bouzyk,M.M., Anderson,K.A.,

Grellier,L.D., Tobin,F.L., Souchet,M. and Brill,A.

JOURNAL Direct Submission

SUBMITTED (02-JUN-1999) Cardiovascular Pharmacology, SmithKline

Beecham, 4 rue du Chesnay-Beauregard, Saint-Gregoire 35760, France

FEATURES

source

Location/Qualifiers
1..2152
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.23-p34.3 between markers DIS434 and DIS507"
75..587
/note="25 kDa heat stress protein; cHSP; preferentially
expressed in cardiovascular and insulin-sensitive tissues;
interacts with C-terminal domain of alpha-filamin or
actin-binding-280, ABP-280"
/product="cardiovascular heat shock protein"
/protein_id="AA020022.1"
/db_xref="GI:6635998"
/translation="MSHRTSSTFRARSHSSSSSSSTSSASRALPAQDPPEKA
LSMFSDPDSFMRPHEPAPFAPGAGNITLDDPAFAVDVDFSEDTITTSN
NHIEVAERKLADGVTMTFAHKQLPEDVDPTSVTSALREDGSLTIRARRPHEHV
QDFFREIKI"
2155..2120

polysignal

ORIGIN

Alignment Scores:
Pred. No.: 1,048-50 Length: 2152
Score: 817.00 Matches: 150
Percent Similarity: 98.05% Conservative: 1
Best Local Similarity: 97.40% Mismatches: 1
Query Match: 96.34% Indels: 2
DB: 9 Gaps: 0

US-09-977-418-8 (1-152) x AFI55908 (1-2152)

QY 1 MetProHisLeuTyrTlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20

DB 1201 ATGCCACATCTGATATAGATGGGGTTTTCACATACAGCTGGTTGTAAGAAAACCTGAT 1260

QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGly 40

DB 1261 GAAATCCCTGCGCTCTGCGCTGCTGCGGCTCCACAGGCGCAAGCGGGGTGGGG 1320

QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 1321 TGGGGCTGGCTCTTCCCTCCACAGGCTGTGTTCTTGGGCTCTCCATCGACAGACA 1380

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrrp 80

DB 1381 GGATCACCTTAACAGATGGAAGCCAGGCGATGATGGGGCTTTGGGTCCTCGAGGTGG 1440

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLeu 100

DB 1441 ACCCCAGCTTCTTGCCACCTTCCCTCCGCGAGTCAGGCTCTCATCCATCCCTCTTT 1500

QY 100 uileTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120

DB 1501 ATCTAAGATCTATAGCTCGGTGGTGTGTAACAACAACCCCTATCGTGTCTTCAAT 1560

QY 120 snThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140

DB 1561 ATACTCGCATTCACATGCTGTGAGGCCAATTCCAGAGCTTTCATCAATCAGATTACAA 1620

QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152

DB 1621 TCTCATTTTCATTAAACGGGAAACATCCCGACGACC 1658

RESULT 3 HSA243191 2152 bp mRNA linear PRI 14-DEC-1999

LOCUS Homo sapiens mRNA for cardiovascular heart shock protein.

ACCESSION HSA243191.1 GI:6580425

VERSION HSA243191.1 GI:6580425

KEYWORDS cHSP gene; heat shock protein.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Krief, S., Faivre, J. F., Robert, P., le Douarin, B., Brument-Larignon, N., Lefriere, I., Bouzyk, M. M., Anderson, K. M., Geller, L. D., Tobin, F. L., Souchet, M. and Bril, A.
Identification and characterization of cvhsp, A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues

JOURNAL J. Biol. Chem. 274 (51), 36592-36600 (1999)

MEDLINE 20062883
PUBMED 10593960
2 (bases 1 to 2152)

REFERENCE Krief, S.
Direct Submission
Submitted (17-JUN-1999) Krief, S., Cardiovascular Pharmacology, Smtkline Beecham, 4 rue du Chesnay Beauregard - Saint-Gregoire, 35760, FRANCE

FEATURES
source
1. 2152
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="heart"
/dev_stage="fetal"
1. 2152
/gene="cvhsp"
75..587
/gene="cvhsp"
/function="small heat stress protein"
/codon_start=1
/product="heat shock protein"
/protein_id="CA63258.1"
/db_xref="GI:6580426"
/db_xref="GOA:Q9UB99"
/db_xref="Swiss-Prot:Q9UB99"
/translation="MSHRTSSTFRAERSFHSSSSSSSTSSASRALPAQDPMEKA LMSDFGFSFMRPHSEPLAPAPGAGNITGLDYEFAVDVDFSEDIVITSN NHIEVRAEKLAADGVMTVFAHKQCLPEVDVPSVTSALREDSLTIRARHPHTHV CQPTREIKT"
polYA_signal
2115..2120
/gene="cvhsp"

ORIGIN
polYA_signal
2115..2120
/gene="cvhsp"

Alignment Scores:
Pred. No.: 1.04e-50 Length: 2152
Score: 817.00 Matches: 150
Percent Similarity: 98.05% Conservative: 1
Best Local Similarity: 97.40% Mismatches: 1
Query Match: 96.34% Indels: 2
DB: 9 Gaps: 0

US-09-977-418-8 (1-152) x HSA243191 (1-2152)

QY 1 MetProHsLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
DB 1201 ATGCCAATCTGATATAGATGGGGTTTTCCAATACAGCTGCTCGATTAACATGCAT 1260

QY 21 GluThrProAlaValleuArgLeuLeuGlyProProGlyLysAlaLysTPGlyTPGly 40
DB 1261 GAAACTCTGCCGCTCCGCGCTCTCGGGGCTCCAGGCAAGCCAACTGGGGTGGGG 1220

QY 41 TTPGlyTPPSePheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1221 TGGGGCTGGCTCTCTCCCTCCACAGCCCTGTTCTTGGGGCTGCTCCACAGCAGACA 1280

QY 61 GlySerProAnaGAspGlySerGlnGlyMetAspGlyAlaLeuGlyProAspGlyTTP 80
DB 1281 GGATCCTTAACAGATGAGACCAAGGCAATGATGGGGCTTTGGGTTCTCGAGGTTGG 1440

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100

DB 1441 ACCCAAGCTTCTGGCACCCTTCCCTCCGGGAGAGACACTCTCCATCCATCCCTCTTT 1500

QY 100 vtleTygIuseerlleGlySerValCysValThr-ThHisProTyArgCysProSerA 120
DB 1501 AATCTATGATATATAGAGCTCGTGTGTGTACACACACACCCCATGATGTTGCTTCA 1560

QY 120 snThGlnHisTyHisTrpLeuArgProAnSerGluLeuSerGlnIleArgPheThrI 140
DB 1561 ATACTAGCATTCATTCACATGGTGTGAGGCCAATTCAGACTTTCTCAATACATTACA 1620

QY 140 lseSerllePheIleAsnGlyGlnThrSerProSerHis 152
DB 1621 TCTCATTTTCATTAACGGGAAACATCCCGAGCCAC 1658

RESULT 4
HSM603488
LOCUS HSM603488 2747 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp686p1416 (from clone DKFZp686p1416).
ACCESSION AL832181
VERSION AL832181.1 GI:21732726
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2747)
Wambutt, R., Heubner, D., Mewes, H. W., Weil, B., Aml, C., Osanger, A., Pofo, G., Han, M. and Wiemann, S.
Direct Submission
Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr. 1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp686p1416) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
source
1. 2747
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686p1416"
/tissue_type="human cervix"
/clone_lib="686 (synonym: h1cc3). Vector pSPori1_Sfi; host DH10B; sites SfiI + SfiIb"
/dev_stage="adult"
polYA_signal
2626..2631
polYA_site
2641

ORIGIN
polYA_signal
2641

Alignment Scores:
Pred. No.: 1.33e-50 Length: 2747
Score: 817.00 Matches: 150
Percent Similarity: 98.05% Conservative: 1
Best Local Similarity: 97.40% Mismatches: 1
Query Match: 96.34% Indels: 2
DB: 9 Gaps: 0

US-09-977-418-8 (1-152) x HSM603488 (1-2747)

QY 1 MetProHsLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
DB 1711 ATGCCAATCTGATATAGATGGGGTTTTCCAATACAGCTGCTCGATTAACATGCAT 1770

QY 21 GluThrProAlaValleuArgLeuLeuGlyProProGlyLysAlaLysTPGlyTPGly 40
DB 1771 GAAACTCTGCCGCTCTCGGGCTCTCGGGGCTCCAGGCAAGCCAACTGGGGTGGGG 1830

QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 1831 TGGGGGTGGCTCTTCCCTCCACAGGCGCTGTGTCTTGGGGCTGCTCCCATGACACA 1890
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 1891 GGATTCACCTTACAGAGATGGAAGCCAGGCGCATGATGGGGCTTTGGGCTCTGAGGTTGG 1950
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerile 100
 Db 1951 ACCCGAGCTTCTTGGCACCTTCCCTCCGGGAGTACGCTCCATCCATCCCTCTCTT 2010
 QY 100 uileTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
 Db 2011 AATCTATGANTCTATAGCTCGCTGGTGTGTACACACACACACCCCTATCTTCTTCAA 2070
 QY 120 snThrGlnHisTyrHisTTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
 Db 2071 ATACTCAGCATTTCCATGTGTGAGGCCAATTCAGAGCTTCTCAATCAGATTACAA 2130
 QY 140 IeSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 2131 TCTCCATTTCATTAAAGGGAACATCCCGAGCCAC 2168
 RESULT 5
 AL355994/c 109377 bp DNA linear PRI 10-JUN-2003
 LOCUS Human DNA sequence from clone RP11-5P18 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL355994 GI:31616669
 VERSION AL355994.11
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 109377)
 AUTHORS Harrison, E.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jun 11, 2003 this sequence version replaced gi:31407704.
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chp1 RP11-5P18 is from the library RPCT-11.1 constructed by the group of

Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.
 location/Qualifiers
 1..109377
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-5P18"
 /clone_1fb="RPCT-11.1"

ORIGIN

Alignment Scores:
 Pred. NO.: 5.55e-49 Length: 109377
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 1
 Query Match: 96.34% Indels: 2
 DB: Gaps: 0

US-09-977-418-8 (1-152) x AL355994 (1-109377)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluTysLeuHis 20
 Db 46193 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATCAGCTGCTGTGATAACTGCAAT 46134
 QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaTyrTTPGlyTTPGly 40
 Db 46133 GAAATCTCTCCCTCCCTCCCTGCTGGGCTCTCCAGGCGCAAGCGGAGGGGTGGGG 46074
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 46073 TGGGGGTGGCTCTTCCCTCCACAGGCGCTGTGTCTTGGGGCTGCTCCCATGACACA 46014
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 46013 GGATTCACCTTACAGAGATGGAAGCCAGGCGCATGATGGGGCTTTGGGCTCTGAGGTTGG 45954
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerile 100
 Db 45953 ACCCGAGCTTCTTGGCACCTTCCCTCCGGGAGTACGCTCCATCCATCCCTCTCTT 45894
 QY 100 uileTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
 Db 45893 AATCTATGANTCTATAGCTCGCTGGTGTGTACACACACACACCCCTATCTTCTTCAA 45834
 QY 120 snThrGlnHisTyrHisTTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
 Db 45833 ATACTCAGCATTTCCATGTGTGAGGCCAATTCAGAGCTTCTCAATCAGATTACAA 45774
 QY 140 IeSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 45773 TCTCCATTTCATTAAAGGGAACATCCCGAGCCAC 45736

RESULT 6

LOCUS CO721116 2138 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 7050 from Patent WO2068579.
 ACCESSION CO721116
 VERSION CO721116.1 GI:42281973
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
 Kites, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 Patent: WO 02068579-A 7050 06-SEP-2002;
 JOURNAL PE Corporation (NY) (US)

```

FEATURES
  source
    Location/Qualifiers
      1..2138
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
  Pred. No.:      2,84e-50      Length:      2138
  Score:          811.00        Matches:      149
  Percent Similarity: 97.40%    Conservative: 1
  Best Local Similarity: 96.75%  Mismatches: 2
  Query Match:    95.64%       Indels:      2
  DB:              6           Gaps:        0

US-09-977-418-8 (1-152) x CQ721116 (1-2138)

QY      1  MePProHsLeuTYrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db      1203 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTCTGATTAACATGCAT 1262
QY      21  GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
Db      1263 GAAACTCTGCGCGCTCCGCGCTGCTGGGGCTCCAGGCAAGGCCACGCGGGTGGGGG 1322
QY      41  TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      1323 TGGGGCTGGTCTCTCCCTCCACAGCCCTGtGtCTTGGGGCTGCTCCCATGCGAGACA 1382
QY      61  GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db      1383 GGATCACCCTACAGATGAGATGAGAGCCAGGGCATGATGGGGCTTGGGCTCCGAGGTGG 1442
QY      81  ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db      1443 ACCCCAGCTTCTTGCCACCTTCCCTCCGGGCACTGCTCCATCCATCCCTCTTT 1502
QY      100  uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
Db      1503 AATTATGAAATCTATAGGCTCGGTGTGTATACACACACCCTATCGTGTCTTCA 1562
QY      120  snThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
Db      1563 ATTACTCGACATTACCATTTGTTGAGGCCAATTCCAGAGCTTTCTCAATACAGATTACAA 1622
QY      140  leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db      1623 TCTCCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 1660

RESULT 7
AX017502      2192 bp      DNA      linear      PAT 07-SEP-2000
LOCUS      Human nucleic acid sequence originating in normal mammary tissue.
DEFINITION      AX017502
ACCESSION      AX017502.1
VERSION      AX017502.1
KEYWORDS      GI:10042299
ORGANISM      Homo sapiens (human)
SOURCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1
      Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
      Piarasky,C.
      Human nucleic acid sequences from normal breast tissue
      Patent: WO 9947655-A 48 23-SEP-1999;
      SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
      BERND (DE); ROSENTHAL ANDRE (DE); HINZMANN GES FUER GENOMFORSCHUN
      (DE); PIARASKY CHRISTIAN (DE)
FEATURES
  source
    Location/Qualifiers
      1..2192
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
  Pred. No.:      2,91e-50      Length:      2192
  Score:          811.00        Matches:      149
  Percent Similarity: 97.40%    Conservative: 1
  Best Local Similarity: 96.75%  Mismatches: 2
  Query Match:    95.64%       Indels:      2
  DB:              6           Gaps:        0

US-09-977-418-8 (1-152) x AX017502 (1-2192)

QY      1  MePProHsLeuTYrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db      1198 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTCTGATTAACATGCAT 1257
QY      21  GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
Db      1258 GAAACTCTGCGCGCTCCGCGCTGCTGGGGCTCCAGGCAAGGCCACGCGGGTGGGGG 1317
QY      41  TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      1318 TGGGGCTGGTCTCTCCCTCCACAGCCCTGtGtCTTGGGGCTGCTCCCATGCGAGACA 1377
QY      61  GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db      1378 GGATCACCCTACAGATGAGATGAGAGCCAGGGCATGATGGGGCTTGGGCTCCGAGGTGG 1437
QY      81  ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db      1438 ACCCCAGCTTCTTGCCACCTTCCCTCCGGGCACTGCTCCATCCATCCCTCTTT 1497
QY      100  uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
Db      1498 AATTATGAAATCTATAGGCTCGGTGTGTATACACACACCCTATCGTGTCTTCA 1557
QY      120  snThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
Db      1558 ATTACTCGACATTACCATTTGTTGAGGCCAATTCCAGAGCTTTCTCAATACAGATTACAA 1617
QY      140  leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db      1618 TCTCCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 1655

RESULT 8
BD135192      2192 bp      DNA      linear      PAT 18-SEP-2002
LOCUS      Human nucleic acid sequence originating in normal mammary tissue.
DEFINITION      BD135192
ACCESSION      BD135192.1
VERSION      BD135192.1
KEYWORDS      GI:23230137
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens (human)
SOURCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1
      Speitt,T., Hinzmann,B., Schmitt,A., Piarasky,C., Duhl,E. and
      Rosenthal,A.
      Human nucleic acid sequence originating in normal mammary tissue
      Patent: JP 2002506639-A 39 05-MAR-2002;
      METAGEN GESELLSCHAFT FUER GENOME FORSCHUNG MBH
COMMENT      OS Homo sapiens (human)
      PN JP 2002506639-A/39
      PD 03-MAR-2002
      PF 19-MAR-1999 JP 2000536838
      PR 20-MAR-1998 DE 198 13 835.0
      PI THOMAS SPEITT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PIARASKI,
      PI EDGAR DUHL,
      PI ANDRE ROSENTHAL
      PC C12N15/09,A61K48/00,A61P43/00,A61P43/00,C07K14/47,
      PC C07K16/18,
      PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68//A61K38/00, PC
      C12N15/00.

```

```

PC C12N5/00/A61K37/02
CC Human nucleic acid sequence originating in normal mammary CC
FH Key tissue Location/Qualifiers
FT source 1..2192 /organism="Homo sapiens (human)"
FEATURES
    source
        Location/Qualifiers
            1..2192
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,91e-50 Length: 2192
Score: 811.00 Matches: 149
Percent Similarity: 97.40% Conservative: 1
Best Local Similarity: 96.75% Mismatches: 2
Query Match: 95.64% Indels: 2
DB: Gaps: 0
US-09-977-418-8 (1-152) x BD135192 (1-2192)
QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGluIleuHis 20
Db 1198 ATCCCATCTGTATTAATGAGGATTTTCCATACAGCTGTCGATTAACCTGCAT 1257
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaAlaIleTyrGly 40
Db 1258 GAAACCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCT 1317
QY 41 ThrGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyValAlaAlaProIleGlnThr 60
Db 1318 TGGGGCTGGCTCTTCCCTCCCAAGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1377
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 1378 GATACACCTTAACAGATGAGAGCCAGGACATGATGAGGCTTGGGCTTGCAGAGTTGG 1437
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaAlaHisProSerProSerIle 100
Db 1438 ACCCGAGCTTCTTGGCACCTTCCCTCCCGGAGAGTCAAGCTCCATCCATCCCTCTTT 1497
QY 100 uileTyrGlySerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
Db 1498 AATCTATGATCATATGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
QY 120 snThrGlnHisTyrHisTyrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140
Db 1558 ATACTCAGATTACCATTTGTTGAGGCGCAAAATTCAGAGCTTCTCAAAATCAGATTACA 1617
QY 140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1618 TCTCCATTTCATTACGGGGAACATCCCGAGCCAC 1655
RESULT 9
BC006319 2589 bp mRNA linear PRI 29-JUN-2004
LOCUS BC006319
DEFINITION Homo sapiens heat shock 27Kda protein family, member 7
(cardiomyosin), mRNA (cDNA clone MGC:12642 IMAGE:4131506),
complete cds.
ACCESSION BC006319
VERSION BC006319.2 GI:33873500
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2589)
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

```

```

TITLE JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
NHL-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13623438.
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kiong, P., Lario, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantiridis, S., Thomas, P.J., Touchman, J.W.,
Turgerson, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: 1 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 20070250.
FEATURES
    source
        Location/Qualifiers
            1..2589
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="MGC:12642 IMAGE:4131506"
                /tissue_type="Muscle, rhabdomyosarcoma"
                /clone_lib="NIH MGC 17"
                /lab_host="DH10B-R"
                /note="vector: pOTB7"
            1..2589
                /gene="HSPB7"
                /note="synonym: CVHSP"
                /db_xref="LocusID:27129"
                /gene="HSPB7"
                /codon_start=1
                /product="heat shock 27Kda protein family, member 7
                (cardiomyosin)"
                /protein_id="AA06319.1"
                /db_xref="GI:13623439"
                /db_xref="LocusID:27129"

```

/translation="MSHRTSSTRRAERFHSSSSSSSSSSSSSASRALPAODPPMEKA
LSMSPDPSFMRSEPLAPFAPPGAGNKKTLGDAFEAVYVRPSPEDDITVTSSN
NHTEVRAEKLAADGTVMNTFAHKCOLPEDVDPTSVTSALREDSGLTIRARPHTEHV
QOTFREIKI"

ORIGIN

Alignment Scores:

Pred. No.:	3,44e-50	Length:	2589
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	9	Gaps:	0

US-09-977-418-8 (1-152) x BC006319 (1-2589)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1621 ATGCCACATCTGTATATAGATGGGGTTTTCACATACAGCTGGTTCGTGATTAACATGCAT 1680
QY 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40
Db 1681 GAAACTCTCGCGCTCGCGCTCTCGGGCTTCAGGCAAGCCACGCTGGGGTTGGGG 1740
QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db 1741 TGGGGTGTGCTCTCTCTCCCTCCACAGGCTGTCTTCTGGGGCTGCCATGACAGACA 1800
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
Db 1801 GGATCACCCTAACACAGATGGAAGCAGGAGCATGATGGGGCTTGGGCTCTTCAGAGTTGG 1860
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 1861 ACCCACCTCTCTTCCACCTTCCCTCCGGGAGTACGCTCTCATCATCCCTCTTT 1920
QY 100 uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
Db 1921 AATTAATGAATCTATAGCTCGGTGTGTAAACACACACACCCCTATCGTTCCTTCA 1980
QY 120 snThrGlnHisTyrHisIleTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 1981 ATACTCGCATTAACATGCTGTGAGCCAAATTCAGACCTTTCTCAATCAGATTTCAA 2040
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db 2041 TCTCCATTTTCATTAAACGGGAAACATCCCGAGCCAC 2078

RESULT 10
LOCUS CQ783843 2701 bp DNA linear PAT 17-MAR-2004
DEFINITION Sequence 3983 from Patent EP1396543.
ACCESSION CQ783843
VERSION CQ783843.1 GI:45503747
KEYWORDS

SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

TITLE Primers for synthesizing full length cDNA clones and their use
JOURNAL Patent: EP 1396543-A 3983 10-MAR-2004;
Research Association for Biotechnology (JP)
FEATURES
SOURCE 1..2701
Location/Qualifiers

FEATURES
SOURCE 1..2701
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
792..1151
/note="unnamed protein product"

/codon_start=1
/protein_id="CAF86888.1"
/db_xref="GI:45503748"
/translation="MQGLHSLTAHPETAPPGAGNKKTLGDAFEAVYVRPSPED
IIVTSSNHIEVRAEKLAADGTVMNTFAHKCOLPEDVDPTSVTSALREDSGLTIRAR
PHTEHVQOTFREIKI"

ORIGIN

Alignment Scores:

Pred. No.:	3.6e-50	Length:	2701
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	6	Gaps:	0

US-09-977-418-8 (1-152) x CQ783843 (1-2701)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1766 ATGCCACATCTGTATATAGATGGGGTTTTCACATACAGCTGGTTCGTGATTAACATGCAT 1825
QY 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40
Db 1826 GAAACTCTCGCGCTCGCGCTCTCGGGCTTCAGGCAAGCCACGCTGGGGTTGGGG 1885
QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db 1886 TGGGGTGTGCTCTCTCTCCCTCCACAGGCTGTCTTGGGGCTCTCCATGACAGACA 1945
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
Db 1946 GGATCACCCTAACACAGATGGAAGCAGGAGCATGATGGGGCTTGGGCTCTTCAGAGTTGG 2005
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 2006 ACCCACCTCTCTTCCACCTTCCCTCCGGGAGTACGCTCTCATCATCCCTCTTT 2065
QY 100 uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
Db 2066 AATCTATGAATCTATAGCTCGGTGTGTAAACACACACACCCCTATCGTTCCTTCA 2125
QY 120 snThrGlnHisTyrHisIleTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 2126 ATACTCGCATTAACATGCTGTGAGCCAAATTCAGACCTTTCTCAATCAGATTTCAA 2185
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db 2186 TCTCCATTTTCATTAAACGGGAAACATCCCGAGCCAC 2223

RESULT 11
LOCUS BD127861 2701 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127861
VERSION BD127861.1 GI:23222806
KEYWORDS JP 2002017375-A/3292.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

TITLE Primers for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 3292 22-JAN-2002;
HELIIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PS JP 2002017375-A/3292
PD 22-JAN-2002 JP 20020251172
PF 07-JUL-2000 JP 20020251172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO

COMMENT
TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO

FEATURES	source	FT	CDS	Location/Qualifiers
ORIGIN				Location/Qualifiers (792)..(1148).
Alignment Scores:				1..2701
Pred. No.:	3.6e-50	Length:	2701	/organism="Homo sapiens"
Score:	811.00	Matches:	149	/mol_type="genomic DNA"
Percent Similarity:	97.40%	Conservative:	1	/db_xref="taxon:9606"
Best Local Similarity:	96.75%	Mismatches:	2	
Query Match:	95.64%	Indels:	2	
DB:	6	Gaps:	0	
US-09-977-418-8 (1-152) x BD127861 (1-2701)				
QY	1	MetProHisLeuTyrIleAspGlyAlaPheProIleGlnLeuValArgGluIysLeuHis	20	
DB	1766	ATGCCACATCTGTATATAGATGGGTTTTCATACACAGCTGGTGTGTATTAACCTGCAT	1822	
QY	21	GIutHrProAlaValIleuArgLeuLeuGlyProProGlyIysAlaLysTyrGlyTyrGly	40	
DB	1826	GAATCTCTCGCGCTCGCTCGCTCGGGCCCTCCAGGCAAGGCGACCTGGGGTTGGGGGG	1885	
QY	41	TyrGlyTyrPheSerPheSerLeuProGlnIleAlaCysVal***GlyAlaAlaProMetGlnThr	60	
DB	1886	TGGGCTGTGCTCTTCTCCCTCCACAGGCTGTGTCTTGGGGCTGTCCCATGGACAA	1945	
QY	61	GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr	80	
DB	1946	GGATCAGCTTACAGAGATGGAAAGCCAGGGGCAATGATGGGGCTTGGTCTCGAGCTTGG	2005	
QY	81	ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerIle	100	
DB	2006	ACCCGAGCTTCTTGGCACCTTCCCTCCGGGAGACAGCACTCCATCCATCCCTCTCTT	2065	
QY	100	uIeTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA	120	
DB	2066	AATCATTAATCTAATAGCTGGGTGTGTGTAAACACACACACCCCATGTCTCTTCAA	2122	
QY	120	snrHrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI	140	
DB	2126	ATATCACACATTAACATATGTTGTAGAGCCAAATTCAGAGCTTCTCAATACAGATTACA	2185	
QY	140	IeSerIlePheIleAsnGlyGlnTyrSerProSerHis	162	
DB	2186	TCTCCATTTTCATTACCGGGGAAACATCCCGAGGCTAC	2223	
RESULT 12				
AK075151	2701 bp	mRNA	linear	PRI 03-SEP-2002
LOCUS	Homo sapiens cDNA FLJ90670.1.f1.s clone PLACR1005539,			similar to
DEFINITION	ACTIN POLYMERIZATION INHIBITOR.			
ACCESSION	AK075151			
VERSION	AK075151.1	GI:22761054		
KEYWORDS	oligo capping; f1s (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

AUTHORS
 Isegai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahara,K., Masuno,Y., Oki,T., Okano,K., Yoshikawa,Y.,
 Kotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Nimomiyu,K.
TITLE
 NEDO human cDNA sequencing project
JOURNAL
 Unpublished
REFERENCE
 2 (bases 1 to 2701)
AUTHORS
 Isegai,T. and Otsuki,T.
TITLE
 Direct Submission
JOURNAL
 Submitted (25-MAR-2002) Takao Isegai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yama, Kisarazu, Chiba 252-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology; cDNA library construction:
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center, cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
FEATURES
 source
 1..2701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACB1005539"
 /issue_type="placenta"
 /clone_id="PLACB1"
 /note="Cloning vector: pME18SFL3"
ORIGIN
 Alignment Scores:
 Pred. No.: 3.6e-50 Length: 2701
 Score: 811.00 Matches: 149
 Percent Similarity: 97.40% Conservative: 1
 Best Local Similarity: 96.75% Mismatches: 2
 Query Match: 95.64% Indels: 2
 DB: Gaps: 0
 US-09-977-418-8 (1-152) x AK075151 (1-2701)
 Qy 1 MeepProhisLeuYrYrIleAspGlyValPheProIleGlnLeuValArgGlyLeuHis 20
 Db 1766 ATGCCAATCTGTRATATAGATGGGGTTTCCATATAGCTGTCTGTGATAAATCGAT 1825
 Qy 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGlyTTPGly 40
 Db 1826 GAAACTCTTCGCCGCTCGCCCTGCTGGGGCTCCACAGGCAAGGCCACGTGGGTGGGG 1885
 Qy 41 TTPGlyTTPsePseSerLeuProGlnAlaCysVal**GlyAlaAlaPromeGlnThr 60
 Db 1886 TGGGGCTGGTCTCTCTCCCTCCACAGGCGCTGAGTTCTTGGGGCTGCTCCATGCAACA 1945
 Qy 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 1946 GGATCACTTAACAGAGATGAGACAGGAGCGATGGATGGGCTTGGGGTCTCCGAGTTGG 2005
 Qy 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 2006 ACCCCAGCTTCCTGGCCACCTTCCTCCCTCGGGCAGTCAAGCTCTCATTCATCCCTCTTT 2065
 Qy 100 uileTyrGluSerIleGlySerValCysValIthr-ThrHisProTyrArgCysProserA 120
 Db 2066 AATTATGAATCTATAGGCTCGGTGTGTGAACACACACACCCTATCGTTCCTTCAA 2125
 Qy 120 snThrglnHisTyrHisTTPLeuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140
 Db 2126 ATATCTCAGCATTCACCATGTGTTAGAGCCAATTCAGAGCTTTCTCAAAATCAGATTTCACA 2185
 Qy 140 leserIlePheIleasnGlyGlnThrseProserHis 152
 Db 2186 TCTCCATTTCATTACGGGAAACATCCCCGAGCCAC 2223

RESULT 13
AK095911
LOCUS 2142 bp mRNA linear PRI 30-JAN-2004
DEFINITION Homo sapiens cDNA FLJ38592 fls, clone HEART1000185, highly similar
to Homo sapiens mRNA for cardiovascular heat shock protein.
AK095911
VERSION AK095911.1 GI:21755262
KEYWORDS cDNA capping; fls (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Makatsutsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawabata, Y., Isono, Y., Nakamura, Y.,
Nagatani, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsura, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirada, S., Chiba, Y., Ishida, S., Oho, Y., Takiguchi, S., Watanabe, S.,
Yoshida, M., Hattori, T., Kusanagi, Y., Kanehori, K., Takahashi, F.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuma, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Sichohata, N., Sano, S., Moriya, S., Momiya, H., Satoh, N., Takam, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kimaaga, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Oho, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, T., Okamoto, S., Okitani, R., Kawakami, T.,
Naguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T.,
Oyama, H., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegaki, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)

JOURNAL
PUBMED 14702039
REFERENCE
AUTHORS 2
Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M.,
Shimizu, F., Wakebe, H., Ota, T., Hishigaki, H., Watanabe, T., Ozaki, K.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makatsutsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,
Takahashi, F., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isegaki, T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2142)
Isegaki, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isegaki, PIJ Project (HRI Team), 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5' & 3' end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers

source 1. 2142
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEART1000185"
/tissue="heart"
/clone_id="HEART1"
/note="Cloning vector: pME18FLU3"

Alignment Scores:
Pred. No.: 5,586-50 Length: 2142
Score: 807.00 Matches: 148
Percent Similarity: 97.40% Conservative: 2
Best Local Similarity: 96.10% Mismatches: 2
Query Match: 95.17% Indels: 2
DB: 9 Gaps: 0

US-09-977-418-8 (1-152) x AK095911 (1-2142)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 1208 ATGCCACATCTGTATATAGATGGGGTCTTTCATATACAGCTGGTTCGATTAACCTGAT 1267
QY 21 GlnTrpProAlaValLeuAlaArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
DB 1268 GAACTCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 1327
QY 41 TrrGlyTrrPsePheSerLeuLeuProGlnAlaCysVal**GlyAlaAlaPromGlnThr 60
DB 1328 TGGGGCTGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1387
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLysGlyProArgGlyTyr 80
DB 1388 GGATCACTCAACAGAGATGGAAAGCCAGGCGATGATGGCGCTTGGCTTCACAGTTGG 1447
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 1448 ACCCCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1507
QY 100 uLeuTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
DB 1508 AATCTATGATATTAAGGCTCGTGTGTGTACACACACACCCCTTATGCTTGTCTTCA 1567
QY 120 snrThGlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
DB 1568 ATACTACACATTAACCATGTTGAGGCGCAATTCAGACTTTCATCAATCAGATTTCACA 1627
QY 140 leSerIlePheIleAsnGlyGlnTrpSerProSerHis 152
DB 1628 TTCATATTTTCAATTAACGGGGAACATCCCGAGCCAC 1665

RESULT 14
HSDJ336M4/c
LOCUS 76241 bp DNA linear PRI 05-JUN-2003
DEFINITION Human DNA sequence from clone Rpl-336M4 on chromosome 1p36.11-36.2
Contains the 5' UTR of the gene MIZ-1, pseudogenes similar to part
of Carboxypeptidase D, the oncogene and the 3' end of the gene for a cardiovascular
heat shock protein. Contains ESTs, STSs, GSSs and two putative CpG
islands, complete sequence.
AL096775
VERSION AL096775.10 GI:7529574
KEYWORDS HMG; CpG islands; heat shock protein; MIZ-1; Tre; USF6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 76241)
Bird, C.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 9, 2000 this sequence version replaced gi:6002162.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Ch1
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP3-336M4 is from the library RPci-3 constructed by the group of
Pier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone
RP3-336M4. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP3-336M4 is at 76241 in this sequence.
The true right end of clone RP1-134019 is at 100 in this sequence.

FEATURES

```
source
1..76241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.11-36.2"
/clone="RP3-336M4"
/clone.lib="RPci-3"

repeat_region
2..174
/note="AluYb8 repeat: matches 1..174 of consensus"
complement(183..559)

misc_feature
/note="match: STS: Em:G37055 Em:N75818"
609..935
/note="match: STS: Em:AL117303 Em:HS442J24S"
629..912
/note="match: GSS: Em:AQ264259"
634..852
/note="match: GSS: Em:B74942"
634..833
/note="match: GSS: Em:AQ311353"
634..768
/note="match: GSS: Em:AQ317636"
641..888
/note="match: GSS: Em:AQ678277"
complement(648..859)
/note="match: STS: Em:G17164"
648..813
/note="match: STS: Em:HS108ME1"
complement(648..785)
/note="match: GSS: Em:AQ687510"
650..927
/note="match: GSS: Em:AQ69729"

misc_feature
551..859
/note="match: GSS: Em:AQ754783"
652..877
/note="match: GSS: Em:AQ282607"
complement(654..814)
/note="match: GSS: Em:AQ118318"
658..935
/note="match: GSS: Em:AQ376808"
661..912
/note="match: STS: Em:G40890"
complement(664..931)
/note="match: GSS: Em:AQ455403"
complement(676..877)
/note="match: GSS: Em:AQ933387"
679..898
/note="match: STS: Em:G01739"
680..884
/note="match: GSS: Em:AQ589116"
complement(680..877)
/note="match: GSS: Em:AQ23541"
684..889
/note="match: GSS: Em:AQ378981"
687..850
/note="match: GSS: Em:AQ538621"
688..893
/note="match: GSS: Em:AQ550571"
complement(688..819)
/note="match: GSS: Em:AQ337044"
690..908
/note="match: GSS: Em:AQ20968"
690..850
/note="match: GSS: Em:AQ341026"
695..833
/note="match: STS: Em:G33512 Em:HS1045WH9"
complement(708..934)
/note="match: GSS: Em:AQ77227"
complement(708..834)
/note="match: GSS: Em:AQ064770"
708..831
/note="match: GSS: Em:AQ387275"
complement(728..936)
/note="match: STS: Em:AQ268772"
complement(745..911)
/note="match: GSS: Em:B64703"
complement(746..864)
/note="match: GSS: Em:AQ148978"
776..938
/note="match: GSS: Em:AQ984702"
789..922
/note="match: GSS: Em:AQ198586"
789..911
/note="match: GSS: Em:AQ30058"
1025..1133
/note="match: STS: Em:AQ30058"
1496..1581
/note="MIR repeat: matches 100..226 of consensus"
1850..2106
/note="L2 repeat: matches 2613..2709 of consensus"
complement(join(2349..2435,5293..5409))
/gene="dJ336M4.1"
/gene="dJ336M4.1"
/product="dJ336M4.1 (Miz-1)"
/note="match: CDNA8: Em:U14556 Em:U23396 Em:Y09723"
match: ESTs: Em:AQ21389 Em:U31841 Em:A155871 Em:AA624099
Em:AA78691 Em:AQ18620 Em:AA408296 Em:AA294814
Em:AA432021 Em:AA481026"
/evidence="not_experimental"
2539..2582
/note="L2 repeat: matches 2648..2692 of consensus"
2583..2682
/note="L2 repeat: matches 2576..2680 of consensus"
2842..3213
repeat_region
```

```
/note="L2 repeat: matches 1731. .2159 of consensus"
3218. .3530
/note="AluYo repeat: matches 1. .312 of consensus"
repeat_region 3568. .3783
/note="AluCo repeat: matches 1. .196 of consensus"
repeat_region 3793. .3874
/note="L2 repeat: matches 2181. .2268 of consensus"
repeat_region 3883. .4185
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 4490. .4709
/note="MIR repeat: matches 17. .260 of consensus"
4718. .5487
/note="CpG island"
/evidence=not_experimental
5044. .5255
/note="106 copies 2 mer cc 58% conserved"
repeat_region 5576. .5794
/note="MIR repeat: matches 31. .262 of consensus"
repeat_region 5925. .6236
/note="L2 repeat: matches 373. .693 of consensus"
repeat_region 6321. .6632
/note="AluY repeat: matches 1. .309 of consensus"
6646. .6771
/note="PRIM_A repeat: matches 1. .133 of consensus"
6808. .6930
/note="L1PB2 repeat: matches 6032. .6155 of consensus"
6931. .7240
/note="AluY repeat: matches 1. .310 of consensus"
7241. .7930
/note="L1PB2 repeat: matches 5337. .6032 of consensus"
7931. .8240
/note="AluSx repeat: matches 1. .301 of consensus"
8241. .8966
/note="L1PB2 repeat: matches 4628. .5337 of consensus"
9024. .9084
repeat_region

Alignment Scores:
Pred. No.: 2.07e-48 Length: 76241
Score: 807.00 Matches: 148
Percent Similarity: 97.40% Conservative: 2
Best Local Similarity: 96.10% Mismatches: 2
Query Match: 95.17% Indels: 2
DB: 9 Gaps: 0

US-09-977-418-8 (1-152) x HSDU336M4 (1-76241)
QY 1 MetProHisLeuTyriLeaSpGlyValPheProIleGlnLeuValArgGlyLeuHis 20
44299 ATGCCACATCTGATATAGATGAGGGGTTTTCATACAGCTGGTGGTGAATAAGCAT 44240
DB
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaLeuTyrGlyTyrPgly 40
44239 GAAACTCTCGCGCTCGCGCTGCTGGGCGCTCCAGGCAAGGCGACGTGGGGTGGGG 44180
DB
QY 41 TrpGlyTyrPsePheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
44179 TGGGGCTGCTCTCTCCCTCCCAAGGCGTGTCTTGGGGGTGCTCCAGCAGACA 44120
DB
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
44119 GGATCACCATTACAGAGATGAGAGCAGGCGATGATGGGCTTTGGGCTCTCAAGTTGG 44060
DB
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100
44059 ACCCCAGCTTCTTCCACCTTCCCTCCGGGAGCTAGCTCTCATCATCAATCCCTCTT 44000
DB
QY 100 uileTyGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
43999 AATCTATGAATCTATAGGCTCGGTGTGTATACACACACACCCCTATCGTGTCTTCAA 43940
DB
QY 120 snTrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
43939 ATACTCAGCATTTACCATTTGGTGTGGGCGCAATTTCAAGCTTTTCAATCAGATTTTACA 43880
DB
```

```
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
43879 TCTCATTTTCATTACCGGGAACATCCCGACGAC 43842
DB
RESULT 15
AX747447 2410 bp mRNA linear PAT 20-JUN-2003
LOCUS AX747447
DEFINITION Sequence 972 from Patent EP1308459.
ACCESSION AX747447
VERSION AX747447.1 GI:32131835
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Isogai,T., Sugiyama,T., Otsuki,T., Makamatsu,A., Sato,H., Ishii,S.,
Yamamoto,O.,I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,W., Nagahara,K. and
Masuko,Y.
Full-length cDNA sequences
Patent: EP 1308459-A 972 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES
source
location/Qualifiers
1..2410
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1.46e-49 Length: 2410
Score: 802.00 Matches: 148
Percent Similarity: 96.75% Conservative: 1
Best Local Similarity: 96.10% Mismatches: 3
Query Match: 94.58% Indels: 2
DB: 6 Gaps: 0

US-09-977-418-8 (1-152) x AX747447 (1-2410)
QY 1 MetProHisLeuTyriLeaSpGlyValPheProIleGlnLeuValArgGlyLeuHis 20
1478 ATGCCACATCTGATATAGATGAGGGGTTTTCATACAGCTGGTGGTGAATAAGCAT 1537
DB
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaLeuTyrGlyTyrPgly 40
1538 GAAACTCTCGCGCTCGCGCTGCTGGGCGCTCCAGGCAAGGCGACGTGGGGTGGGG 1597
DB
QY 41 TrpGlyTyrPsePheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
1598 TGGGGCTGCTCTTCTCCCTCCCAAGGCGTGTCTTGGGGGTGCTCCAGTGAACA 1657
DB
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
1658 GGATCACCATTACAGAGATGAGAGCAGGCGATGATGGGCTTTGGGCTCTCAAGTTGG 1717
DB
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100
1718 ACCCCAGCTTCTTCCACCTTCCCTCCGGGAGCTAGCTCTCATCATCAATCCCTCTT 1777
DB
QY 100 uileTyGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
1778 AATCTATGAATCTATAGGCTCGGTGTGTATACACACACACCCCTATCGTGTCTTCAA 1837
DB
QY 120 snTrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
1838 ATACTCAGCATTTACCATTTGGTGTGGGCGCAATTTCAAGCTTTTCAATCAGATTTTACA 1897
DB
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
1898 TCTCATTTTCATTACCGGGAACATCCCGACGAC 1995
DB
```

Search completed: October 24, 2004, 17:12:47
job time : 3591 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

CM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:04:28 ; Search time 84 Seconds

(without alignments)
1286.189 Million cell updates/sec

Title: US-09-977-418-8
Perfect score: 848
Sequence: 1 MRLPLVDGVPIQLVREKLH.....LSQIRPTISIFINGERSPSH 152

Scoring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 segs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US09977418/runat_22102004_115948_1048/app_query.fasta_1.327
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORB=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=M099977418 @CGN 1.1 69 runat 22102004_115948_1048 -NCPU=6 -ICPU=3
-NMAP -LARGEODURY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGIOG
-DEV.TIMEOUT=120 -WARM.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seg.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seg.*
- 3: /cgn2_6/ptodata/1/ina/5A.COMB.seg.*
- 4: /cgn2_6/ptodata/1/ina/5B.COMB.seg.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seg.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seg.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	11.4	1330	4 US-10-329-668-17	Sequence 17, Appl
2	97	11.4	2000	4 US-08-016-434-1291	Sequence 1291, Ap
3	97	11.4	2000	4 US-09-919-497-10	Sequence 10, Appl
4	95.5	11.3	4403765	3 US-09-103-840A-2	Sequence 2, Appl
5	95.5	11.3	4411529	3 US-09-103-840A-1	Sequence 1, Appl
6	93.5	11.0	17000	4 US-09-691-220-3	Sequence 3, Appl
7	93	11.0	15202	4 US-09-602-787A-341	Sequence 341, Appl
8	93	11.0	15202	4 US-08-922-635-21	Sequence 21, Appl
9	92	10.8	13857	4 US-08-620-312D-75	Sequence 75, Appl
10	92	10.8	18609	3 US-08-943-731-1	Sequence 71, Appl
11	91.5	10.8	3992	4 US-09-944-807-9	Sequence 9, Appl
12	91	10.7	2733	4 US-09-799-451-480	Sequence 480, App

C 13	91	10.7	4480	3 US-09-191-171-7	Sequence 7, Appl
C 14	91	10.7	4480	3 US-09-385-707-7	Sequence 7, Appl
C 15	91	10.7	6238	4 US-09-638-696C-6	Sequence 6, Appl
C 16	91	10.7	8460	1 US-08-469-005A-9	Sequence 9, Appl
C 17	90	10.6	1512	3 US-08-955-918C-8	Sequence 8, Appl
C 18	90	10.6	1512	3 US-08-697-766A-8	Sequence 8, Appl
C 19	90	10.6	1785	4 US-08-264-578-7	Sequence 7, Appl
C 20	90	10.6	1801	3 US-08-955-918C-6	Sequence 6, Appl
C 21	90	10.6	1801	3 US-08-697-766A-6	Sequence 6, Appl
C 22	90	10.6	8519	3 US-09-261-907-1	Sequence 1, Appl
C 23	89.5	10.6	1218	3 US-09-012-072-1	Sequence 1, Appl
C 24	89.5	10.6	1218	3 US-09-120-601-1	Sequence 1, Appl
C 25	89.5	10.6	2205	1 US-08-081-610-1	Sequence 1, Appl
C 26	89	10.5	1719	4 US-09-016-444-1472	Sequence 1472, Ap
C 27	89	10.5	1734	4 US-08-948-559A-7	Sequence 7, Appl
C 28	89	10.5	1734	2 US-08-663-808-1	Sequence 1, Appl
C 29	89	10.5	1734	2 US-09-188-469-7	Sequence 7, Appl
C 30	89	10.5	1734	2 US-09-332-740-1	Sequence 1, Appl
C 31	89	10.5	1734	3 US-09-188-466-1	Sequence 7, Appl
C 32	89	10.5	1734	3 US-09-397-238A-7	Sequence 7, Appl
C 33	89	10.5	1734	3 US-09-368-282-1	Sequence 1, Appl
C 34	89	10.5	1734	4 US-09-566-708A-1	Sequence 1, Appl
C 35	89	10.5	2326	2 US-08-231-193A-41	Sequence 41, Appl
C 36	89	10.5	2326	2 US-08-486-273A-41	Sequence 41, Appl
C 37	89	10.5	2326	3 US-08-486-273A-41	Sequence 41, Appl
C 38	89	10.5	2326	3 US-08-940-066A-41	Sequence 41, Appl
C 39	89	10.5	2326	3 US-08-940-035A-41	Sequence 41, Appl
C 40	89	10.5	2326	3 US-08-935-105A-41	Sequence 41, Appl
C 41	89	10.5	2326	4 US-09-648-797-41	Sequence 41, Appl
C 42	89	10.5	2326	4 US-09-386-123-41	Sequence 41, Appl
C 43	89	10.5	3243	2 US-08-231-193A-44	Sequence 44, Appl
C 44	89	10.5	3243	2 US-08-486-273A-44	Sequence 44, Appl
C 45	89	10.5	3243	3 US-08-480-474-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-10-329-668-17
; Sequence 17, Application US/10329668
; Patent No. 6696473
; GENERAL INFORMATION:
; APPLICANT: Martin Richard
; APPLICANT: Brennon Todd
; APPLICANT: Kahl Jeffrey Dean
; APPLICANT: Wang Tie-Lin
; TITLE OF INVENTION: HETEROCYCLIC MODULATORS OF NUCLEAR RECEPTORS
; FILE REFERENCE: 38205-3001
; CURRENT APPLICATION NUMBER: US/10/329,668
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/342,720
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(837)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank XM_042579
; DATABASE ENTRY DATE: 2002-02-06
; US-10-329-668-17

Alignment Scores:
Pred. No.: 2.38
Score: 97.00
Percent Similarity: 35.54%
Best Local Similarity: 26.51%
Query Match: 11.44%
DB: 4
Length: 1330
Matches: 44
Conservative: 15
Mismatches: 51
Indels: 56
Gaps: 10

US-09-977-418-8 (1-152) x US-10-329-668-17 (1-1330)

QY 2 ProHleuLeuTylleaspolyValPheProIleGlnLeuValArgGluLysLeuHis--- 20
 Db 396 CCTTACTACTCTTGCCTGCGGA-----CAACAAGACTGCACAGT 434

QY 21 ---GluThrProAla-----ValLeuArgLeuLeuGlyProProGlyLys--- 34
 Db 435 GGACAGGCCCGACCGGACCGCTGTACAGTACGCGCTTACAGAGTGCCTGCGCACAG 494

QY 35 -----AlaLysTyrGlyTyr 39
 Db 495 CATGAGAGGAGGCGGCTACAGAGAGCGCTACGCGGGAAGACAGAGATGGGGAGTGG 554

QY 40 GlyTyrGlyTyrSerPheSerLeuProGlnAlaCys-----Val**GlyAlaAla 56
 Db 555 GGAAGGGGCTGGGGAGGCCCGGAGAGATGCTGTGACAGAGATCTGGAGGCGAGACT 614

QY 57 PrometGlnThrGlySer---ProAsnArgAspGlySerGlnGlyMetAspGly----- 73
 Db 615 TGCTGTGAACAGAAAGATGACACAGGCGCTTGAGGGGTCTGAGGGGACCGGGGGTACGG 674

QY 74 -----AlaLeuGlyProArgGlyTyrThrProAla 83
 Db 675 CAGCAGCGTGAAGTGGGGTCAATCCACTCTCTCGTATGGGGGTTGGGGAGGAG 734

QY 84 SerCySHleuProLeuArgGlnSerAlaLeuHisProSerProSerLeuIleTyrGlu 103
 Db 735 TCT-----AGTCTGTCTTACATCCCTCCCTCTCTTCCCTCAT 776

QY 104 SerIleGlySerValCysValThrThrHisProTyrArgCys-----ProSerAsnThr 121
 Db 777 AACCTTCCTAAC-----ACTACTTGGAGATGAGAGTGTGCAAAACAGAGCTTTTC 827

QY 122 GlnHisTyrHisTyrLeu 127
 Db 828 AACCATCTGAGGTGATG 845

RESULT 2
 US-09-016-434-1291
 ; Sequence 1291, Application US/09016434
 ; Patent No. 6500938
 ; GENERAL INFORMATION:
 ; APPLICANT: Janice Au-Young
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 ; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 ; NUMBER OF SEQUENCES: 1490
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 PORTER DRIVE
 ; CITY: PALO ALTO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/016,434
 ; FILING DATE: HERWITH
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Zeller, Karen J.
 ; REGISTRATION NUMBER: 37,071
 ; REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (650) 855-0555
 ; TELEFAX: (650) 845-4166
 ; INFORMATION FOR SEQ ID NO: 1291:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2000 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GENBANK
 ; CLONE: g29980
 ; US-09-016-434-1291

Alignment Scores:
 Pred. No.: 4.2 Length: 2000
 Score: 97.00 Matches: 37
 Percent Similarity: 40.46% Conservative: 16
 Best Local Similarity: 28.24% Mismatches: 42
 Query Match: 11.44% Indels: 36
 Gaps: 10

US-09-977-418-8 (1-152) x US-09-016-434-1291 (1-2000)

QY 27 ArgLeuLeuGlyProProGlyLysAlaLysTyrGly-----TyrGlyTyr 41
 Db 184 AGACAAAGACACCGGACCGGCGCATCACGCTGGGGCGCTGAGCGCGGCATGTCATG 243

QY 42 GlyTyrSer---PheSerLeuPro-----Gln 49
 Db 244 AAGTGGGACCCCTGCAGAGCTGAGGCGCTGCGGCGCTGCGGGGAGCGAGCGGCGCAT 303

QY -50 AlaCysVal**GlyAlaAlaProMetGlnThrGlySerPro----- 63
 Db 304 GCTTGTGCTGAGACTCCGCTCTCTCTTCAACCGCCGACATGCGCGCTG 363

QY 64 AsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArg-----GlyTyr 80
 Db 364 TCAAGTGCCTTCAACACCATGCTGCGGCGC-----GGCCAAAGGCGGCATGCGCTGG 420

QY 81 ThrProAlaSerCySHleuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 421 AGC---ACATCGTCCCAAGCGGAGCTCGGCGCGCTGCTGCGGCGCGCTACACAG 477

QY 101 IleTyrGlnSerIleGlySerValCysValThrThr-----HisProTyrArgCysPro 118
 Db 478 CCGTGG-----TGTTCCTGAGACGAGCGGCGCGCGCTGACGCGCCA 522

QY 119 SerAsnThrGlnHisTyrHisTyrLeuArgPro 129
 Db 523 AGCGCGACGCGACCCGCGCGCTGG---CGGCGG 552

RESULT 3
 US-09-919-497-10
 ; Sequence 10, Application US/09919497
 ; Patent No. 6773883
 ; GENERAL INFORMATION:
 ; APPLICANT: Mutter, George L.
 ; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
 ; FILE REFERENCE: B0801/725
 ; CURRENT APPLICATION NUMBER: US/09/919,497
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/221,735
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 100
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 2000
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-919-497-10

Alignment Scores:

Pred. No.:	4.2	Length:	2700
Score:	97.00	Matches:	3000
Percent Similarity:	40.46%	Conservative:	16
Best Local Similarity:	28.24%	Mismatches:	42
Query Match:	11.44%	Indels:	36
DB:	4	Gaps:	10

US-09-977-418-8 (1-152) X US-09-919-497-10 (1-2000)

QY	27	ArgLeuLeuAllyProProGlyValAlaLysTrpGly-----	TrpGlyTrp	41
Db	184	AGCAAAAGAGACACCCGAGGCGCCATACGCTGGGGCCCTAGAGCGGCATGGTCATGG		243
QY	42	GlyTyrSer---PheSerLeuPro-----	Gln	49
Db	244	AAGTGGGACCCCTGACGCTGAGCGCCTGGGGCCGTGCTGGAGGAGCGAGCGGCACAT		303
QY	50	AlaCysVal**GlyAlaAlaProMetGlnThrGlySerPro-----		63
Db	304	GCCGCGCTGAGACTCGCTCCGCTCTTTCGCTTTAAAGCGCGGCACATCGCGGCTGTG		363
QY	64	AsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArg-----	GlyTyr	80
Db	364	TCAACGCGCCCTTCAGACCATGTGTGGCGCC---GGGCAAGGGCCCATGGGCTGTG		420
QY	81	ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu		100
Db	421	AGC---ACATGTGTGCCCAACGCCGAGACTCGGGGGCGCTGTGGCGGGCGGTACACAG		477
QY	101	IleTyrGlnSerIleLeuGlySerValCysValThrTyr-----HisProTyrArgCysPro		118
Db	478	CCGTGG-----TGTGGCTGAAGAGCGAGCGCCGCTGTGACGGCGCCA		522
QY	119	SerAsnThrGlnHisTyrHisTrpLeuArgPro		129
Db	523	AGCGCAGCGCACCTGTGGCTGTGG---CGGCGG		552

RESULT 4

```

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Parent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

US-09-977-418-8 (1-152) X US-09-103-840A-2 (1-4403765)

[illegible]

RESULT 5

US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Parent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

US-09-977-418-8 (1-152) X US-09-103-840A-1 (1-4411529)

```

OY 30 |Y|P|P|R|G|I|Y|Y|A|L|A|Y|T|P|G|I|Y|T|P|G|I|Y|T|P|S|R|P|H|E|L|A|P|R|G|I|N 49
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754022 GGGCGCTCTGGCCCTGGCGCGCTGGTCGCTTTCTGTCTGTGCTGTCTGTCTGCGC-----CGC 753965E
OY 50 |A|C|G|Y|A|L|***G|Y|A|L|A|A|P|R|O|M|E|T|G|I|N|H|T|G|I|Y|S|R|P|R|O|A|N|A|G|A|S|P|G|I|Y|S|E|G|I|N 69
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753968 CACCTGGCGC-----AAAGATTCATCTTACGCGC 753942A
OY 70 |G|I|Y|E|A|S|P|G|I|Y|A|L|A|L|E|U|-----G|I|Y|P|R|O|A|G|I|Y|T|I|P|H|P|R|O 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753941 GGGTGTGCTCTGGAGTGTCTGTGATCGACGCTGGAGTCGGCTGGCGCGCCCTGTGTGTATACCA 753882Z

```



```

; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19940764.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940765.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940766.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940830.0
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940831.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940832.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940833.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941395.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942077.7
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942078.5
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942079.3
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942088.2
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 341
; LENGTH: 1986
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1963)
; OTHER INFORMATION: RXA00878
US-09-602-787A-341

```

```

Alignment Scores:
Pred. No.: 10.3      Length: 1986
Score: 93.00      Matches: 40
Percent Similarity: 32.93%      Conservative: 14
Best Local Similarity: 24.39%      Mismatches: 55
Query Match: 10.97%      Indels: 56
DB: 4      Gaps: 5

```

```

US-09-977-418-8 (1-152) x US-09-602-787A-341 (1-1986)
QY 15 ValArgLysLeuHsGluThrProAlaValLeuArgLeuGlyProGlyLys 34
    :::::|||||:::|||||:::|||||:::|||||
Db 796 TTACGGAAACAGTACGATTAACCGCGCCGCAATCCGTTGGTGGACCGCATGATAC 855
QY 35 AlAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGlnAlaCysVal**Gly 54
    :::::|||||:::|||||:::|||||:::|||||
Db 856 TCTCCGGGGGCGC--GGCACTAAATA-TCACTCTCCGGCCATTCACCTGATTTCT 911
QY 55 AlAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGlnAlaCysVal**Gly 54
    :::::|||||:::|||||:::|||||:::|||||
Db 912 ACCGCACTTGCACGCGCATTTCAACATCGCGCATGATCA-----TGCATCAAGC 965
QY 75 LeuGlyPro-----ArgGlyTrpThrProAlaSerCysHisLeu 77
    :::::|||||:::|||||:::|||||:::|||||
Db 966 TTGACCATGAGCGCAGCAACCGGTATCATGATGACTACATGTCGACTCCGCGAG 1025
QY 78 -----ArgGlyTrpThrProAlaSerCysHisLeu 87
    :::::|||||:::|||||:::|||||:::|||||
Db 1026 GCGCATCGCGCGGTCCCGGCACTCTTGCAGTCAAGCGCAGAGTCCAGCCACCG 1085
QY 88 ProLeuArgLysLeuHsGluThrProAlaValLeuArgLeuGlyProGlyLys 107
    :::::|||||:::|||||:::|||||:::|||||

```

```

Db 1086 CCAATAGAGCTTGTGACGCCACCTCGCTCAGACATCCACGATCA----- 1136
QY 108 ValCysValThrThrHisProTyrArgCysProSerAsnThrGlnHisTrpLeu 127
    :::::|||||:::|||||:::|||||:::|||||
Db 1137 -----GTCTGGGACAGCGCCCGCGGCTCCCGCTGGAGAACACGATCG----- 1181
QY 128 ArgProAsnSerGlnLeuSerGlnLeuArgPheThrIleSerIlePheIleArgGlyGlu 147
    :::::|||||:::|||||:::|||||:::|||||
Db 1182 -----AATTCAAA 1190
QY 148 ThrSerProSer 151
    :::::|||||:::|||||:::|||||:::|||||
Db 1191 AGCTACCTTCG 1202

```

```

RESULT 8
US-08-922-635-21
; Sequence 21, Application US/08922635A
; Patent No. 6033871
; GENERAL INFORMATION:
; APPLICANT: PILLETZ, John E.
; APPLICANT: IVANOV, Tina R.
; TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
; TITLE OF INVENTION: AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: Corrected Sequence Listing
; Patent No. 6033871
; CURRENT APPLICATION NUMBER: US/08/922,635A
; EARLIER FILING DATE: 1997-09-03
; EARLIER APPLICATION NUMBER: 08/650,766
; EARLIER FILING DATE: 1996-05-20
; EARLIER APPLICATION NUMBER: 60/012,600
; EARLIER FILING DATE: 1996-03-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 15202
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-922-635-21

```

```

Alignment Scores:
Pred. No.: 174      Length: 15202
Score: 93.00      Matches: 38
Percent Similarity: 41.44%      Conservative: 8
Best Local Similarity: 34.23%      Mismatches: 32
Query Match: 10.97%      Indels: 34
DB: 3      Gaps: 7

```

```

US-09-977-418-8 (1-152) x US-08-922-635-21 (1-15202)
QY 25 ValLeuArgLeuLeuGlyPro-----ProGlyLysAlaLysTrp---GlyTrpGlyTrp 41
    :::::|||||:::|||||:::|||||:::|||||
Db 1532 ATCTTGCTCTTATGTGAGACACACAGCCCTGCTGGGCCCTGCTGGGCTGGGGTTGG 1591
QY 42 GlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAla-AlaProMetGlnThrG1 61
    :::::|||||:::|||||:::|||||:::|||||
Db 1592 GGG-----ACAGGTGCGACGACCTGCTTCACAG 1621
QY 61 YserProAsnArgAspGlySerGlnGlyMetAspGlyLysLeuGlyProArgGly----- 79
    :::::|||||:::|||||:::|||||:::|||||
Db 1622 GGTCAACACAGAGGAGGCGATGCTT-CTCAGAGCT-----GGTCTCCGCGGGGAGC 1674
QY 80 -----TyrThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLe 94
    :::::|||||:::|||||:::|||||:::|||||
Db 1675 ACATGCGAGGGGTGCTGCTGATGCGACGCTGTGCTTGGTGAAGACCTCCCAATT 1734
QY 94 u-----HisProSerProSerLeuIleTyrGlnSerIleGlySerValCysVa 110
    :::::|||||:::|||||:::|||||:::|||||
Db 1735 GCTCTGATGCCACATCAAGCTCT-----CTAGAGACCGGAGGT 1776
QY 110 ThrThrHisProTyrArgCysProSerAsn 120
    :::::|||||:::|||||:::|||||:::|||||
Db 1777 GTCTGACAGCGCCCTGAGGCTCCCTCTGATC 1807

```


NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-707-7

Alignment Scores:
Pred. No.: 50.3 Length: 4480
Score: 91.00 Matches: 32
Percent Similarity: 40.82% Conservative: 8
Best Local Similarity: 32.65% Mismatches: 34
Query Match: 10.73% Indels: 25
DB: 3 Gaps: 5

US-09-977-418-8 (1-152) x US-09-385-707-7 (1-4480)

QY 30 GYProProGlyLysAlaLysTPGly-----TTP----- 39
DB 1908 GGTCCGCTGCTGCTTGTAAATGGGGGTGCCGGAAGTGGGGAAGCTCCGATCT 1849
QY 40 -----GlyTTPGlyTTPserPheSerLeuProGlnAla 50
DB 1848 GCTGCGGACGACCTTCTCTGCTCCAGATGAGATGAGTGGCGGACCT----- 1795
QY 51 CysVal**GlyAlaAlaProMetGlnThrglySerProAsnArgAspGlySerGlnGly 70
DB 1794 ---GCGGGCGGGGGCGCTTCCAGACACCGAGGTTCAGCG---CGGCCCGGGGACAGCG- 1742
QY 71 MetAspGlyAlaLeuGlyProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg 90
DB 1741 GCTGAGGGGCGCA---GAACACCGGGGCGGACGAGGGGCGGCGGACCTTCTCTGTGAG 1685
QY 91 GlnSerAlaLeuHisProSerProSerLeuLeuTyrGlnSerIleGlySerVal 108
DB 1684 GGAAGATGATGTCAGCGCGACGCCCGCTCCAGTGAAGAGTGTGATCCGTC 1631

RESULT 15
US-09-639-696C-6/c
Sequence 6, Application US/09639696C
Patent No. 6524835

GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
NELSON, Paul V.
CLEMENS, Peter R.
MORRIS, Charles P.
HOPWOOD, John J.

TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON PEAHODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 2249/304
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-832-7555
TELEFAX: 516-832-7555
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6238 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-639-696C-6

Alignment Scores:
Pred. No.: 79.7 Length: 6238
Score: 91.00 Matches: 32
Percent Similarity: 40.82% Conservative: 8
Best Local Similarity: 32.65% Mismatches: 34
Query Match: 10.73% Indels: 25
DB: 4 Gaps: 5

US-09-977-418-8 (1-152) x US-09-639-696C-6 (1-6238)

QY 30 GYProProGlyLysAlaLysTPGly-----TTP----- 39
DB 3666 GGTCCGCTGCTGCTTGTAAATGGGGGTGCCGGAAGTGGGGAAGCTCCGATCT 3607
QY 40 -----GlyTTPGlyTTPserPheSerLeuProGlnAla 50
DB 3606 GCTGCGGACGACCTTCTCTGCTCCAGATGAGATGAGTGGCGGACCT----- 3553
QY 51 CysVal**GlyAlaAlaProMetGlnThrglySerProAsnArgAspGlySerGlnGly 70
DB 3552 ---GCGGGCGGGGGCGCTTCCAGACACCGAGGTTCAGCG---CGGCCCGGGGACAGCG- 3500
QY 71 MetAspGlyAlaLeuGlyProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg 90
DB 3499 GCTGAGGGGCGCA---GAACACCGGGGCGGACGAGGGGCGGCGGACCTTCTGTGAG 3443
QY 91 GlnSerAlaLeuHisProSerProSerLeuLeuTyrGlnSerIleGlySerVal 108
DB 3442 GGAAGATGATGTCAGCGCGACGCCCGCTCCAGTGAAGAGTGTGATCCGTC 3389

Search completed: October 24, 2004, 18:24:49
Job time: 1649 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:01:52 ; Search time 427 Seconds

(without alignments)
1868.646 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPELYIDGVFPIQLVREKHL.....LSQIRFTISIFINGERSPSH 152

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1h
-O=/cgm2_1/USFTO.spool/US09977418/runat_22102004_115947_1008/app_query.fasta_1.327
-DBn_genseq_23Sep04 -QFMT=faclap -SUFFIX=eng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=b1ts -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human4.0.cdi
-LIST=45 -DOCALLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ent -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09977418 @CGN 1.1 470 @runat_22102004_115947_1008 -NCPU=6 -ICPU=3
-NO MAMP -LARGEJOBY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DT TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Genseq_23Sep04:*
1: genseq1980s:*
2: genseq1990s:*
3: genseq2000s:*
4: genseq2001as:*
5: genseq2001bs:*
6: genseq2002as:*
7: genseq2002bs:*
8: genseq2003as:*
9: genseq2003bs:*
10: genseq2003cs:*
11: genseq2003ds:*
12: genseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	846	99.8	1987	4	AAF27852	AAf27852 Human NOV
2	846	99.8	1987	9	ACD40257	ACd40257 Human bre
3	846	99.8	1987	11	ADMS6372	ADms6372 Human CDN
4	846	99.8	1987	12	ADf66734	Adf66734 CDNA enco
5	846	99.8	1987	12	ADl19771	Adl19771 Human NOV
6	846	99.8	1987	12	ADo60244	ADo60244 Human NOV

7	817	96.3	2142	4	AA160565	AA160565 Human pol
8	817	96.3	2142	10	ADCC31919	ADcc31919 Human nov
9	817	96.3	2152	12	ADQ17578	Adq17578 Human sof
10	817	96.3	4056	5	ADM20171	Adm20171 Alternati
11	811	95.6	2192	2	AA233583	Aa233583 Human bre
12	811	95.6	2220	5	ADM19231	Adm19231 Novel hum
13	811	95.6	2701	4	AAK94832	Aak94832 Human ful
14	811	95.6	2701	12	ADL31950	Adl31950 Full leng
15	811	95.6	2783	10	ADCC29984	ADcc29984 Human nov
16	811	95.6	3490	8	ACC46692	Acc46692 Human dlt
17	811	95.6	3503	12	ADQ22267	Adq22267 Human sof
18	802	94.6	2410	10	ADB62818	Adb62818 Human CDN
19	644	75.9	400	9	ACH16935	ACH16935 Human adu
20	578	68.2	1550	4	AAFS4984	Aaf54984 Nucleotid
21	512	68.4	305	9	AAE27801	Aae27801 Human NOV
22	512	68.4	305	11	ADMS6439	Adms6439 Human CDN
23	512	60.4	305	12	ADf66802	Adf66802 Novel hum
24	512	60.4	305	12	ADf66802	Adf66802 Novel hum
25	512	60.4	305	12	ADf66802	Adf66802 Novel hum
26	502	59.2	305	12	ADl19839	Adl19839 Human NOV
27	113.5	13.4	2347	6	ABA93765	Ab93765 Human tra
28	110.5	13.0	4427	5	AA879292	Aa879292 DNA encod
29	107	12.6	175737	6	ABK83571	Abk83571 Human CDN
30	107	12.6	175737	10	ADL13596	Adl13596 Osteoarth
31	107	12.6	175737	12	ADQ18934	Adq18934 Human sof
32	103.5	12.2	21010	4	AAK89247	Aak89247 Human dig
33	103.5	12.2	21010	4	AA105888	Aa105888 Human rep
34	103.5	12.2	21010	4	AB198452	Ab198452 Human tes
35	103.5	12.2	21010	8	AB274628	Ab274628 Secreted
36	103.5	12.2	21010	10	ABZ68148	Abz68148 Human sec
37	103.5	12.2	21024	4	AAK89248	Aak89248 Human dig
38	103.5	12.2	21024	4	AA105889	Aa105889 Human rep
39	103.5	12.2	21024	4	AB198453	Ab198453 Human tes
40	103.5	12.2	21024	8	AB274629	Ab274629 Secreted
41	103.5	12.2	21024	10	ABZ68149	Abz68149 Human sec
42	103	12.1	1701	12	ACH87372	ACH87372 Human gen
43	103	12.1	2888	11	ADM02718	Adm02718 Human CDN
44	103	12.1	2954	4	AA197913	Aa197913 Human neu
45	103	12.1	2954	4	AA198066	Aa198066 Human neu

ALIGNMENTS

RESULT 1	AAf27852	AAf27852 standard; CDNA; 1987 BP.
XX	AAf27852	
AC	AAf27852	
XX		
DT	30-MAR-2001 (first entry)	
XX		
DE	Human NOV4 CDNA.	
XX		
XX	Human; NOVX; antiinflammatory; cytoprotective; neuroprotective;	
KW	cerebroprotective; immunomodulator; vulnery; vasotrophic; gene therapy;	
KW	hyperplasia; tumour; reteleosis; psoriasis; Dupuytren's contracture;	
KW	diabetes; Rheumatoid arthritis; cerebral oedema; Alzheimer's disease; ss.	
OS	Homo sapiens.	
XX		
PN	WO200075321-A2.	
XX		
PD	14-DEC-2000.	
XX		
PF	01-JUN-2000; 2000WO-US015303.	
XX		
PR	03-JUN-1999; 99US-0137322P.	
PR	16-MAR-2000; 2000US-0189810P.	
PR	22-MAR-2000; 2000US-0191158P.	
PR	30-MAR-2000; 2000US-0193086P.	
PR	31-MAY-2000; 2000US-00137322.	
XX		
PA	(CURA-) CUPRAGEN CORP.	

XX Shimkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI, 2001-102403/11.
 DR P-PSDB; AAB61132.
 XX
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing.
 XX
 PS Claim 8; Page 26-27; 194pp; English.
 XX
 CC The present sequence encodes a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)
 CC
 XX
 SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;
 Alignment Scores:
 Pred. No.: 1,92e-57 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: Gaps: 0
 US-09-977-418-8 (1-152) x AAF27852 (1-1987)
 QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIleuHis 20
 Db 991 ATGCCACATCTGATATTAATGGGGTTTTCCTCAATACAGCTGTCGTAATAACTGCAT 1050
 QY 21 GlnThrProAlaValLeuArgLeuGlnGlyProProGlyValValAlaLeuTyrGlyTyr 40
 Db 1051 GAAACTCTGCGCTCTGCGGCTGCTGCGGCTCCAGGCAAGGCAAGGCGGCTGGGGG 1110
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal***GlyValAlaLeuPheGlnThr 60
 Db 1111 TGGGGCTGATCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGCTCCATGCAGACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
 Db 1171 GGGTACCTTAACAGATGGAAGCCAGGCAATGAGTGGGCTTTGGGTCTCGAGTTGG 1230
 QY 81 ThrProAlaSerCysHisLeuProLeuAlaGlnSerAlaLeuHisProSerProSerLeu 100
 Db 1231 ACCCCAGCTTCTTCCACCTTCCCTCCGAGTCAAGCTCTCCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 Db 1291 ATTATGAACTCTTAGGCTCGGTGTGTGAACAACACACCTTAATGTTGCTTCAAT 1350
 QY 121 ThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
 Db 1351 ACTCACCATTTACATGGTTGAGGCGCAATTCAGAGCTTTCATCAATCAGATTTTCAATC 1410
 QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 1411 TCCATTTTCATTAAACGGGGAACATCCCGAGCCAC 1446

RESULT 2
 ID ACD40257 standard; DNA; 1987 BP.
 XX
 AC ACD40257;
 XX
 DT 03-SEP-2003 (first entry)
 XX
 DE Human breast tumour associated protein 47-like polypeptide NOV4 DNA.
 XX
 KM Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
 KM pancreatic cancer; uterine cancer; organ transplantation disorder; ds;
 KM cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
 KM ischaemic heart disease; haemorrhage; peripheral vascular disease; gene;
 KM thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;
 KM tissue regeneration; wound healing; hyperproliferative disorder;
 KM psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
 KM Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KM ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
 KM neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;
 KM gene therapy; epilepsy; breast tumour associated protein 47.
 XX
 OS Homo sapiens.
 XX
 FN US2003027158-A1.
 XX
 PD 06-FEB-2003.
 XX
 PF 15-OCT-2001; 2001US-00977418.
 XX
 FR 03-JUN-1999; 99US-0137322P.
 FR 16-MAR-2000; 2000US-0189810P.
 FR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 XX
 DR WPI; 2003-492028/46.
 DR P-PSDB; ABO23236.
 XX
 PT New nucleic acid sequence encoding a human breast tumor-associated
 PT protein 47-like polypeptide, useful for treating cardiovascular
 PT disorders, neural disorders, diabetes mellitus and cancers.
 XX
 PS Claim 8; Page 16-17; 100pp; English.
 XX
 CC The invention relates to a new isolated NOV4 nucleic acid. The nucleic
 CC acid is useful for identifying a compound that binds the nucleic acid.
 CC The nucleic acid is useful in gene therapy, in screening assays, in
 CC detection assays e.g. chromosomal mapping, cell and tissue typing and
 CC forensic biology, predictive medicine e.g. diagnostic assays, prognostic
 CC assays, monitoring clinical trials, and pharmacogenomics and methods of
 CC treatment including therapeutic and prophylactic. The nucleic acid is
 CC also useful for expressing NOVX protein. The nucleic acid is also useful
 CC to provide polynucleotide reagents e.g. labelled probes that are useful
 CC in an in situ hybridisation technique, for identifying a specific tissue
 CC (for example brain tissue) and for use in forensic science. The nucleic
 CC acid is also useful for mapping genes on a chromosome and thus locating
 CC gene regions associated with genetic disease, identifying an individual
 CC from a minute biological sample and to aid in forensic identification of
 CC biological sample. The nucleic acid is also useful for treating cancer,
 CC especially cancers of the breast, colon, lung, pancreas or uterus, or a
 CC melanoma or sarcoma. The nucleic acid is also useful for treating
 CC disorders related to organ transplantation, cardiovascular diseases,
 CC atherosclerosis, ischaemic heart disease, haemorrhage and systemic lupus
 CC erythematosus. NOVX protein encoded by the nucleic acid is useful for
 CC regulating haematopoiesis, for regeneration of bone, cartilage, tendon

CC ligament and/or nerve tissue growth or regeneration and for wound
CC healing. The nucleic acid is also useful for treating infections,
CC hyperproliferative disorders e.g. psoriasis, and neural disorders
CC including Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral
CC neuropathy, tumours of the nervous system, exposure to neurotoxins, acute
CC brain injury, peripheral nerve trauma or injury and other neuropathies,
CC epilepsy, and/or tremors. The present sequence represents DNA encoding a
CC human breast tumour associated protein 47-like polypeptide
XX

SO Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

Alignment Scores:

Alignment Scores:	
Pred. No.:	1,92e-57
Score:	846.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	99.76%
DB:	
Length:	1987
Matches:	152
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-977-418-8 (1-152) x ACD40257 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValAlpheProIleGlnLeuValArgGlnLeuHis 20
DB 991 ATGCCACATCTGATATAGATGAGGTTTTCATACAGCTGTTCTGTAAACTGCAT 1050
QY 21 GlnThrProAlaValLeuArgLeuGlyProProGlyValAlaYsTyrGlyTyrGly 40
DB 1051 GAAACCTCCGCTCCGCGCCCTGCGGGCTCCAGGAGCAAGTGGGGTGGGG 1110
QY 41 TyrGlyTyrSerHisSerLeuProGlnAlaCysVal**GlyValAlaProMetGlnThr 60
DB 1111 TGGGGCTGGCTCTCCCTCCCAAGGCTGTCGTCGTCGTCCTCCATGACAGACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 1171 GGATCACTTAACAGAGATGAGACCAAGGCGCATGAGTGGGCTTGGGTCTCGAGCTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCCAGCTCTTCCACCTCCCTCCGCGAGCTGCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTACATCTATAGCTCGCTGTGTGAACAACACCCCTATGCTTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTCAGCATATCCATGCTTGGAGCCAAATTCAGACCTTCTCAATTCAGATTACATC 1410
QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 1411 TCCATTTTCATTAACGGGGAACATCCCGAGGCAC 1446

RESULT 3
ADMS6372
ID ADMS6372 standard; cDNA; 1987 BP.
AC ADM56372;
XX
DT 03-JUN-2004 (first entry)
DE
XX Human cDNA encoding cell adhesion molecule NOVA #1.
XX
XX Human; ss; gene; cell adhesion molecule; NOVA; cancer; leukaemia;
KW lymphoma; melanoma; neurological disorder; epilepsy;
KW ischaemic cerebrovascular disease; stroke; Alzheimer's disease;
KW Pick's disease; vesicular transport disorder; cystic fibrosis;
KW diabetes mellitus; Grave's disease; goiter; gastrointestinal disorder;
KW ulcerative colitis; gastric ulcer; duodenal disorder; autoimmune disease;
KW allergic reaction; autoimmune haemolytic anaemia; rheumatoid arthritis;
KW viral infection; bacterial infection; fungal infection;
KW helminthic infection; protozoal infections.

XX Homo sapiens.
OS
XX
XX US2003082554-A1.
PN
XX
XX 01-MAY-2003.
PD
XX
XX 15-OCT-2001; 2001US-00977033.
PF
XX
XX 03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189818P.
PR 22-MAR-2000; 2000US-0191188P.
PR 30-MAR-2000; 2000US-0193086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
XX (CUBA-) CUBAGEN CORP.
PA
XX
XX Shinkets RA, Fernandes E, Heriman J, Vernet C;
PI WPI; 2003-616079/58.
XX
XX P-PsDB; ADMS6371.
DR
XX
XX New nucleic acid encoding human cell adhesion molecule-like proteins,
PT useful for treating e.g. cancers, neurological disorders, viral,
PT bacterial, fungal, helminthic and protozoal infections.
XX
XX Example 3; SEQ ID NO 7; 78pp; English.
XX
XX The invention relates to an isolated nucleic acid encoding a human cell
XX adhesion molecule-like protein, comprising a sequence encoding a
XX polypeptide having a sequence appearing as ADMS6371, a sequence at least
XX 90% identical to the nucleic acid, a sequence encoding a polypeptide
XX having conservative amino acid substitutions to the protein or a fragment
XX comprising at least 20 nucleotides. Also included are an oligonucleotide
XX sequence that is complementary to (and hybridises under stringent
XX conditions with) the nucleic acid (or a portion of it), a vector
XX comprising the nucleic acid, a cell comprising the vector, a
XX pharmaceutical composition comprising the nucleic acid and a
XX pharmaceutical carrier, a process for producing the polypeptide, a
XX process for identifying a compound that binds the nucleic acid, and a
XX compound identified by the process. Disclosed as new are the cDNA and
XX proteins for novel cell adhesion molecules (termed NOVA, being NOVA-23).
XX The NOVA polypeptide, nucleic acid or antibody are useful in the
XX manufacture of a medicament for treating a syndrome associated with a
XX human disease, selected from NOVA-associated disorder, such as cancers
XX (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
XX ovary, testis and uterus), neurological disorders (e.g. epilepsy,
XX ischaemic cerebrovascular disease, stroke, Alzheimer's disease or Pick's
XX disease), disorders of vesicular transport (e.g. cystic fibrosis,
XX diabetes mellitus, Grave's disease, or goiter), gastrointestinal
XX disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
XX autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
XX anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
XX and protozoal infections. The polypeptides can be used as immunogens to
XX produce antibodies and as vaccines. The sequences may further be used in
XX chromosome mapping, identifying individual from minute biological samples
XX (tissue typing), and in forensic identification of a biological sample.
XX NOTE: The authors have mislabelled the sequences as they appear on pages
XX 12-28 of the patent. It is clear from table 3, the examples and the
XX claims that the SEQ ID numbers for the cDNAs should be the odd numbers
XX from 1-45 and the proteins should be the even numbers from 2-46. The
XX present sequence encodes a NOVA cell adhesion molecule of the invention.
XX
XX Sequence 1987 BP; 448 A; 615 C; 529 G; 386 T; 0 U; 9 Other;
SO
XX
XX Alignment Scores:
XX Pred. No.: 1,92e-57
XX Score: 846.00
XX Percent Similarity: 100.00%
XX Best Local Similarity: 100.00%
XX Query Match: 99.76%
XX DB: 11
XX Length: 1987
XX Matches: 152
XX Conservative: 0
XX Mismatches: 0
XX Indels: 0
XX Gaps: 0

US-09-977-418-8 (1-152) x ADM56372 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluLeuHis 20
 Db 991 ATGCCACATCTGTATTAATAGTGGGTTTTCATACAGCTGGTTCGTAAGAACTGAT 1050
 QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyLysAlaLysTyrPgly 40
 Db 1051 GAAACTCTGCGCTGCTGCGCTGCGGAGCTCCAGGCGAAGCCAGTGGGGTTGGGG 1110
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
 Db 1111 TGGGCTGTGCTCTTCTCCCTCCACAGGCTCTGTCTTGGGCTGCTCCCATGACAGCA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1171 GGATCCTACACAGAGATGGAAGCCAGGCGATGATGGGGCTTTGGTCTCGAGTTGG 1230
 QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 1231 ACCCGAGCTTCTTGCCACTTCCCTCCGCGAGTACGCTTCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 Db 1291 ATCTATGATCTATAGGCTGGGTGTGTGTAAACACACACCCTATCGTTGCTTCAT 1350
 QY 121 ThrGlnHisTyrHisTyrLeuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140
 Db 1351 ACTCAGCATTTACCATTTGGTGGAGGCCAATTGAGAGCTTTCCAAATCAGATTACCAATC 1410
 QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 1411 TCCATTTTCATTATACCGGGAACATCCCGAGCCAC 1446
 RESULT 4
 ADF66734
 ID ADF66734 standard; cDNA; 1987 BP.
 XX ADF66734;
 AC ADF66734;
 DT 12-FEB-2004 (first entry)
 DE cDNA encoding novel human protein NOV4.
 XX
 XX cytostatic; hepatotropic; vulnery; antipsoriatic; osteopathic;
 KW antiallergic; antiatherosclerotic; haemostatic; vasotropic;
 KW thrombolytic; antidiabetic; hypotensive; dermatological;
 KW immunosuppressive; antiinflammatory; immunostimulant; fungicide;
 KW viricide; protozoicide; neuroprotective; antineumatic; antiallergic;
 KW antiallergic; antiparkinsonian; nocotropic; anticonvulsant;
 KW NOV4 modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
 KW psoriasis; tissue hypertrophy; osteoarthritis;
 KW atherosclerotic plaque formation; haemorrhage; ischaemic disease;
 KW thrombosis; diabetes mellitus; hypertension; hypothyroidism;
 KW immune deficiency; severe combined immunodeficiency; SCID; infection;
 KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
 KW multiple sclerosis; systemic lupus; erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
 KW autoimmune thyroiditis; insulin dependent diabetes mellitus;
 KW myasthenia gravis; graft-versus-host disease;
 KW autoimmune inflammatory eye disease; asthma; hematopoiesis;
 KW tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
 KW periodontal disease; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
 KW human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX US200319103-A1.
 XX
 PD 23-OCT-2003.
 XX

PF 15-OCT-2001; 2001US-00977639.
 XX
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Fernandes E, Herrman J, Vernet C;
 XX
 DR WPI, 2004-021196/02.
 DR P-PDB; ADF66735.
 XX
 PT Novel substantially NOV4 polypeptide useful for diagnosing, preventing
 PT and treating diseases e.g., cancer, multiple sclerosis, systemic lupus
 PT erythematosus.
 XX
 PS Disclosure; SEQ ID NO 7; 165bp; English.
 XX
 CC The invention describes a substantially purified polypeptide (I) having
 CC amino acid sequence chosen from a fully defined NOV4 sequence (S1) of 708
 CC amino acids as given in the specification, or polypeptide having one or
 CC more conservative amino acid substitutions of (S1), or mutant or variant
 CC of (S1). (I) having (S1) is useful for diagnosing a pathological
 CC condition associated with (I) or its activity in a subject e.g. cancer.
 CC (I) useful in treatment of cancer, hyperproliferative diseases,
 CC cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis,
 CC atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal
 CC disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I)
 CC is useful in treatment of various immune deficiencies and disorders such
 CC as severe combined immunodeficiency (SCID), bacterial infection, viral
 CC infection such as herpes viral infection, protozoan infection such as
 CC malaria, fungal infection such as candidiasis. (I) is also useful in
 CC treating autoimmune disorders such as connective tissue disease, multiple
 CC sclerosis, systemic lupus, erythematosus, rheumatoid arthritis, autoimmune
 CC pulmonary inflammation, Guillain-Barre syndrome, autoimmune
 CC thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis,
 CC graft-versus-host disease and autoimmune inflammatory eye disease and
 CC asthma. (I) useful in regulation of hematopoiesis, regeneration and
 CC tissue growth of bone, cartilage, tendon, ligament and useful for wound
 CC healing and tissue repair. (I) is also useful in treatment of burns,
 CC incisions and ulcers. (I) also useful in treatment of periodontal
 CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective
 CC antitumour and antiinflammatory activity. This sequence encodes a novel
 CC human NOV4 protein.
 CC
 XX
 SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;
 XX
 Alignment Scores:
 Pred. No.: 1,928-57 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 12 Gaps: 0
 US-09-977-418-8 (1-152) x ADF66734 (1-1987)
 QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluLeuHis 20
 Db 991 ATGCCACATCTGTATTAATAGTGGGTTTTCATACAGCTGGTTCGTAAGAACTGAT 1050
 QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyLysAlaLysTyrPgly 40
 Db 1051 GAAACTCTGCGCTGCTGCGCTGCGGAGCTCCAGGCGAAGCCAGTGGGGTTGGGG 1110
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
 Db 1111 TGGGCTGTGCTCTTCTCCCTCCACAGGCTCTGTCTTGGGCTGCTCCCATGACAGCA 1170

QY 61 GlySerProAsnAlaGlySerGlnGlyMetAspGlyAlaLeuGlyProArgIYTP 80
DB 1171 GGATTCACCTAAAGAGATGAGAGCCAGGCGATGATGGGCTTGGCTTCAGAGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCCAGCTCTTGGCAGCTTCCCTCCGACGACGCTCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTAGATCAATCAATAGAGCTCGGTGTGTATCAACAACACCCCATGCTTGTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTyrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTGACGATTACCATGTTGGATGGCCAAATTCAGAGCTTCTCAATCAGATTTCACATC 1410
QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
DB 1411 TCATTTTCATTACCGGGGAAACATCCCGAGCCAC 1446
RESULT 5
AD119771
ID AD119771 standard; DNA; 1987 BP.
XX AD119771;
AC
DT 22-APR-2004 (first entry)
XX
XX Human NOV4 DNA.
XX
XX Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
XX infection; anorexia; cancer; cardiovascular disease; hypertension;
XX atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
XX Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
XX haematopoietic disorder; inflammatory skin disorder; asthma;
XX dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
XX haematopoiesis; wound healing; angiogenesis; chromosome mapping;
XX tissue typing; preventive medicine; pharmacogenomics; gene therapy;
XX anorectic; cardiac; virocid; antibacterial; fungicide; protozoacide;
XX neurotropic; neuroprotective; dermatological; human; gene; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 991..1449
XX FT
XX FT /*tag= b
XX FT /product= "Human NOV protein"
XX FT /transl_except= (pos:1171..1221, aa:Gly-Pro)
XX FT /transl_except= (pos:1144..1152, aa:Val-Gly)
XX FT sig_peptide 991..1152
XX FT /*tag= a
XX FT mat_peptide 1153..1446
XX FT /*tag= c
XX FT /product= "Human mature NOV protein"
XX
XX US2004002134-A1.
XX
XX 01-JAN-2004.
XX
XX 15-OCT-2001; 2001US-00977819.
XX
XX
XX 03-JUN-1999; 99US-0137322P.
XX 16-MAR-2000; 2000US-0189810P.
XX 22-MAR-2000; 2000US-019158P.
XX 30-MAR-2000; 2000US-019308P.
XX 03-MAY-2000; 2000US-0201388P.
XX 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Fernandes ER, Herrman JL, Vernet CM;
XX

DR WPI, 2004-070737/07.
DR P-PSDB; AD119772.
XX
XX New NOVX nucleic acids encoding human KIAA0768 protein-like and human
PT protein PRO-228 polypeptides, useful for treating NOVX-associated
PT disorders.
XX
XX Disclosure: SEQ ID NO 7, 95pp; English.
XX
XX The present invention is based in part on the discovery of novel secreted
CC and membrane-bound polypeptides and their encoding polynucleotides. The
CC nucleic acids and polypeptides are collectively referred as NOVX. The
CC invention is useful for treating, preventing and diagnosing diseases such
CC as metabolic disorders, diabetes, obesity, infectious diseases such as
CC viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
CC cancer, cardiovascular diseases such as hypertension and atherosclerosis,
CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC epilepsy, immune disorders such as osteoarthritis, haematopoietic
CC disorders, inflammatory skin disorders, asthma and various dyslipidemias.
CC The invention is also useful as targets for the identification of small
CC molecules that modulate or inhibit e.g. neurogenesis, cell
CC differentiation, cell proliferation, haematopoiesis, wound healing and
CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
CC typing, preventive medicine and pharmacogenomics. The invention is also
CC useful in gene therapy. The present sequence is human NOV DNA.
XX
SQ Sequence 1987 BP; 448 A; 615 C; 530 G; 386 T; 0 U; 8 Other;
XX
Alignment Scores:
Pred. No.: 1,92e-57 Length: 1987
Score: 846.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 12 Gaps: 0
US-09-977-418-8 (1-152) x AD119771 (1-1987)
QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLeuHis 20
DB 991 ATGCCACATCTGTATATGATGGGGTTTTCATACAGCTGGTTCGTAAGAACTGAT 1050
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyIlyAlaIlyTyrGly 40
DB 1051 GAAACTCTCGCGCTCGCTCGCTGGGGCTCCAGGCAAGGCCAAGTGGGTTGGGG 1110
QY 41 TTPGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1111 TGGGCTGTCTCTCTCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGACACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgIYTP 80
DB 1171 GGATTCACCTAAAGAGATGAGAGCCAGGCGATGATGGGCTTGGCTTCAGAGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCCAGCTCTTGGCAGCTTCCCTCCGACGACGCTCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTAGATCAATCAATAGAGCTCGGTGTGTATCAACAACACCCCATGCTTGTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTyrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTGACGATTACCATGTTGGATGGCCAAATTCAGAGCTTCTCAATCAGATTTCACATC 1410
QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
DB 1411 TCATTTTCATTACCGGGGAAACATCCCGAGCCAC 1446
RESULT 6
ADO60244
ID ADO60244 standard; DNA; 1987 BP.

```

XX AD060244;
XX
XX 15-JUL-2004 (first entry)
XX
XX Human NOV4 DNA.
XX
XX Human; NOVX protein; cancer; hyperproliferative disease; cirrhosis;
XX keloid; psoriasis; tissue hypertrophy; osteoarthritis;
XX atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
XX renal disease; thrombosis; diabetes mellitus; hypertension;
XX hypothyroidism; severe combined immunodeficiency; SCID; infection;
XX malaria; candidiasis; autoimmune disorder; connective tissue disease;
XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
XX autoimmune pulmonary inflammation; Guillain-Barre syndrome;
XX autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
XX autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
XX autoimmune inflammatory disease; Parkinson's disease;
XX Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
XX haematopoiesis; wound healing; tissue repair; antitumor;
XX antiinflammatory; gene; ds.
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX CDS 991..1149
XX /tag= b
XX /product= "NOVX protein"
XX /transl_except= (pos:1147..1149, aa:Xaa)
XX /note= "Xaa corresponds to any amino acid"
XX
XX sig_peptide 991..1065
XX /tag= a
XX mat_peptide 1066..1146
XX /tag= c
XX /product= "Mature NOVX protein"
XX
XX US2003134430-A1.
XX
XX 17-JUL-2003.
XX
XX 15-OCT-2001; 2001US-00977751.
XX
XX 03-JUN-1999; 99US-0137322P.
XX 16-MAR-2000; 2000US-0189810P.
XX 22-MAR-2000; 2000US-0191158P.
XX 30-MAR-2000; 2000US-0193086P.
XX 03-MAY-2000; 2000US-0201388P.
XX 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimketa RA, Fernandes E, Herrman J, Vernet C;
XX
XX MPI, 2004-068928/07.
XX F-PSDB; AD060245.
XX
XX Novel substantially purified NOVX polypeptide for treating severe
XX combined immunodeficiency, candidiasis, cancer, asthma, multiple
XX sclerosis, systemic lupus erythematosus.
XX
XX Disclosure; SEQ ID NO 7; 155pp; English.
XX
XX The invention relates to human NOVX polypeptides and polynucleotides.
XX NOVX sequences are useful in the treatment of cancer, hyperproliferative
XX diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
XX osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
XX heart or renal disease, thrombosis, diabetes mellitus, hypertension,
XX hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
XX Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
XX and disorders such as severe combined immunodeficiency (SCID), bacterial
XX infection, viral infection e.g. herpes viral infection, protozoan
XX infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune

```

```

CC disorders such as connective tissue disease, multiple sclerosis, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease. The invention is useful in
CC regeneration and tissue growth of bone, cartilage, tendon, ligament.
CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
CC the invention also exhibit antitumor and antiinflammatory activities. The
CC present sequence is human NOVX DNA.
XX
XX Sequence 1987 BP; 448 A; 613 C; 532 G; 366 T; 0 U; 8 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,92e-57 Length: 1987
XX Score: 846.00 Matches: 152
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.76% Indels: 0
XX DB: 12 Gaps: 0
XX
XX US-09-977-418-8 (1-152) x AD060244 (1-1987)
XX
XX QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
XX Db 991 ATGCCACATCTGTATATATAGTGGCGTTTTCACATACAGCTGCTGTAATAAATGCAAT 1050
XX
XX QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTrpGly 40
XX Db 1051 GAAACCTCTCCCTCTCTGCGCCCTGCGGCCCTCCAGGCAAGCCAAAGTGGAGTGGGG 1110
XX
XX QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
XX Db 1111 TGGGCGTGGCTCTCTCTCCCTCCACAGGCGCTGTTCTTGGGCGTCTCCATGACAGACA 1170
XX
XX QY 61 GlySerProAsnArgAspLysSerGlnGlyMetLeuAspGlyValLeuGlyProArgGlyTrp 80
XX Db 1171 GGATCACTPAACAGATAGAGAGCCAGCCAGCATGAGAGGGGCTTGGGCTCTGAGAGTTGG 1230
XX
XX QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
XX Db 1231 ACCCCAGCTTCTTCCACCTTCCCTCCCGGAGCTAGCTCTCAATCCATCCCTCTTTTA 1290
XX
XX QY 101 IleTyrGluSerIleGlySerValCysValIThrThrHisProTyrArgCysProSerAsn 120
XX Db 1291 ATCTATGAACTATATAGCTGCTGGGTGTGTATACACACACCCCTATCGTTGTCTTCAAT 1350
XX
XX QY 121 ThrGlnHisTyrHisTrpLeuArgProAsnSerGluLeuSerGlnLeuArgPheThrIle 140
XX Db 1351 ACTCAGCATTAACCATTTGTTGAGGCCAAATTCAGAGCTTCTCAATTCAGATTACAAATC 1410
XX
XX QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
XX Db 1411 TCCATTTTCAATTACGGGGAAACATCCCGAGCCAC 1446
XX
XX RESULT 7
XX AA160565
XX ID AA160565 standard; cDNA; 2142 BP.
XX AC AA160565;
XX
XX DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4554.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX

```

PN WO20015312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-0048725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00596042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang U, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou F, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM41409.
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Claim 1; SEQ ID NO 4554; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (A157798-A161369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral neuropathies, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilization of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 CC
 SQ Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 4.08e-55 Length: 2142
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 1
 Query Match: 96.34% Indels: 2
 DB: 4 Gaps: 0
 US-09-977-418-8 (1-152) x AAI60565 (1-2142)
 QY 1 MetProHisLeuTYrTlleAspGlyValPheProIleGlnLeuValArgGluValSerHis 20
 DB 1198 ATGCCACATCTGATATAGATGGGGTTTTCATTAAGCTGTCGTGATAACTGCAT 1257
 QY 21 GluThrProAlaValAlaLeuArgLeuLeuGlyProProGlyLysAlaIleSTPGLYTPGLY 40
 DB 1258 GAAATCTCTGCCGCTCCGCTCGCTCGGGGCTCCAGGCAAGGCAAGTGGGGTGGGG 1117
 QY 41 TTPGLYTPSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
 DB 1318 TGGGGCTGTGCTCTCTCCCTCCACAGGCTGTGTCCTTGGGGCTGCTCCCATGCACACA 1377
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 1378 GGATCCTTAACAGATGAGAGCCAGGCGCATGATGGGGCTTTGGGCTCTCGAGTTGG 1437

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 1438 ACCCGAGCTTCTTGCACTTCCCTCCGGGAGAGACTCTCCATTCATCCCTCTTT 1497
 QY 100 uileTyGluSerIleGlySerValCysValThr-ThHisProTyrArgCysProSera 120
 DB 1498 AATCTATGATCTATAGGCTCGGTGTGTAAACACACACCCCATGCTGTGCTTCAA 1557
 QY 120 snTHGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
 DB 1558 AATCTCAGCATTTACCATTTGCTTGAAGCCAAATTCAGACTTCTTCAATCAGATTACA 1617
 QY 140 IeSerIlePheIleAsnGlyGluThrSerProSerHis 152
 DB 1618 TCTCATTTTCACTTAACGGGAAACATCCCGAGCCAC 1655
 RESULT 8
 AD31919
 ID AD31919 standard; cDNA; 2142 BP.
 XX
 AC AD31919;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human novel cDNA contig sequence, SEQ ID NO:2001.
 XX
 XX Human; diagnostic; drug screening; forensics; gene mapping;
 XX biodiversity; assessment; Parkinson's disease; Alzheimer's disease;
 XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 XX ulcers; osteoporosis; autoimmune disease; cancer;
 XX molecular weight marker; food supplement; antiparkinsonian; neurotrophic;
 XX neuroprotective; antiischemic; anticoagulant; thrombolytic; vulnerary;
 XX anticancer; osteopathic; immunosuppressive; antiinflammatory; cyostatic;
 XX gene therapy; chromosome 1p36.11-36.2; ss.
 OS Homo sapiens.
 XX
 PN WO2003029271-A2.
 XX
 PD 10-APR-2003.
 XX
 XX 24-SBP-2002; 2002WO-US030474.
 XX
 XX 24-SBP-2001; 2001US-0324631P.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang TY, Zhang Y, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Wang G;
 PI Haley-Vicente D, Drmanac RT;
 XX
 DR WPI: 2003-371981/35.
 DR P-PSDB; AD32686.
 XX
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 XX
 Example 2; SEQ ID NO 2001; 1185bp; English.
 XX
 CC The invention relates to 971 novel human cDNA sequences (AD329919-
 CC AD330889) and the polypeptides they encode (AD330890-AD331860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes

CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (AD017578-AD017579) and the polypeptides encoded by the contigs (AD032628
CC -AD033394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcc_sequences.

XX Sequence 2142 BP; 451 A; 700 C; 569 G; 423 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,08e-55	Length:	2142
Score:	817.00	Matches:	150
Percent Similarity:	98.058	Conservative:	1
Best Local Similarity:	97.408	Mismatches:	1
Query Match:	96.348	Indels:	2
DB:	10	Gaps:	0

US-09-977-418-8 (1-152) x AD031919 (1-2142)

```
QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIysLeuHis 20
DB 1198 ATGCCACATCTGATATATAGATGGGCTTTTCCAAACACCTGCTGATATAACTGCAT 1257
QY 21 GlnThrProAlaValLeuArgLeuGlnGlyProProGlyLysValAlaGlyTyrGly 40
DB 1258 GAAACCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1318 TGGGGCTGGTCTTCTCCCTCCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
QY 61 GlySerProAsnArgSpGlySerGlnGlyMetLeuValAlaLeuGlyProArgGlyTyr 80
DB 1378 GATCACCTTAACAGAGATGAAAGCCAGGCGCATGAGTGGGCTTTGGGCTCTGAGGTTGG 1437
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 1438 ACCCCAGCTTCTGGCACCCTCCCTCCGCGGCAAGTCAGCTCCATCCATCCCTCTTT 1497
QY 100 uilleTrrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
DB 1498 AATCATGATCATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
QY 120 snThrGlnHisTyrHisTrrPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
DB 1558 AATACAGCATTAACCATTTGTTGAGGCCAAATTCAGAGCTTTCTCAATTCAGATTACAA 1617
QY 140 lserTrrPheIleAsnGlyGlnTrrSerProSerHis 152
DB 1618 TCTCCATTTTCATTAACGGGGGAAACATCCCGAGCCAC 1655
RESULT 9
AD017578
ID AD017578 standard; DNA; 2152 BP.
AC AD017578;
XX
XX 26-AUG-2004 (first entry)
```

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 395.
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-UDN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.

XX Example 2; SEQ ID NO 395; 210bp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cyostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.

XX Sequence 2152 BP; 459 A; 700 C; 569 G; 424 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.1e-55	Length:	2152
Score:	817.00	Matches:	150
Percent Similarity:	98.058	Conservative:	1
Best Local Similarity:	97.408	Mismatches:	1
Query Match:	96.348	Indels:	2
DB:	12	Gaps:	0

US-09-977-418-8 (1-152) x AD017578 (1-2152)

```
QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIysLeuHis 20
DB 1201 ATGCCACATCTGATATATAGATGGGCTTTTCCAAATCACCTGCTGATATAACTGCAT 1260
QY 21 GlnThrProAlaValLeuArgLeuGlnGlyProProGlyLysValAlaGlyTyrGly 40
DB 1261 GAAACCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1321 TGGGGCTGGTCTTCTCCCTCCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 61 GlySerProAsnArgSpGlySerGlnGlyMetLeuValAlaLeuGlyProArgGlyTyr 80
DB 1381 GATCACCTTAACAGAGATGAAAGCCAGGCGCATGAGTGGGCTTTGGGCTCTGAGGTTGG 1440
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
```

Dh 1441 ACCCAGCTTCTTGCCACCTTCCCTCCGGGAGTACGCTTCATCCATCCCTCTTT 1500
Qy 100 uUleYrYgluSerlllglySerValThr-ThrHlSPcYrYrAgCYsProSeRa 120
Db 1501 AATCTATGAACTCATAGGCTCGGCTGTGTGTACACACACACCCCTATCGTGTCTTCAA 1560
Qy 120 snThrglnHlStYrHlStYrLeuArgProAsnSerGluSerGlnleArgPheThrI 140
Db 1561 ATATCAGCATTTCCATTTGCTTGAGGCCAATTCCAGGCTTCTCAATACAGATTACAA 1620
Qy 140 lEserllPheleAsnGlyGluThrSerProSerHlS 152
Db 1621 TCTCCATTTCATTAAAGGGGAAACATCCCGAGCCAC 1658
RESULT 10
ADM20171
ID ADM20171 standard; cDNA; 4056 BP.
XX
AC ADM20171;
XX
DT 20-MAY-2004 (first entry)
XX
DE Alternative nucleotide for novel channel/transporter cDNA #28.
XX
KW immunosuppressive; antiarthritis; antirheumatic; antiproliferative;
KW cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW gene therapy; channel/transporter protein; rheumatoid arthritis;
KW neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive; ds.
XX
OS Homo sapiens.
XX
PN W0200154472-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218298P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 06-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0233968P.
PR 12-SEP-2000; 2000US-0233972P.
PR 14-SEP-2000; 2000US-0233988P.
PR 14-SEP-2000; 2000US-0233998P.
PR 14-SEP-2000; 2000US-0234002P.
PR 14-SEP-2000; 2000US-0234012P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239936P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251473P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251907P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM,
XX WPI; 2001-476159/51.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 978; 809pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to an alternative version
CC of the gene of the invention.
XX
SQ Sequence 4056 BP; 800 A; 1323 C; 1182 G; 751 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,48e-55 Length: 4056
Score: 817.00 Matches: 150
Percent Similarity: 98.05% Conservative: 1

Best Local Similarity: 97.40% Mismatches: 1
Query Match: 96.34% Indels: 2
DB: 5 Gaps: 0
us-09-977-418-8 (1-152) x ADM20171 (1-4056)
Qy 1 MetProh1stleuYrllleapglYvalpheprollleglleuvalargulysleuH1s 20
Db 3075 ATGCCACATCTGTATATAGATGGGGTTTTCACATCAGCTGGTTCGTATTAACATGCAT 3134
Qy 21 GluThrProAlaValleuAArgleuLeuGlyProProGlylyAlaIstPglYtPglY 40
Db 3135 GAACCTCCTCCGCGCTCGGCGCTCGGCGCTCCAGGCAAGGCAAGTGGGTTGGGG 3194
Qy 41 TtPglYtPserPheSerleuProGlnAlaCysVal***GlyAlaIalPrometGlnThr 60
Db 3195 TGGGGCTGCTCTCTCCCTCCCAAGGCTGTGTCTGGGGCTCTCCATGACAGACA 3254
Qy 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaIleuGlyProArgGlyTTP 80
Db 3255 GGATCACCCTTAACAGAGATGGAACCGGCGATGATGGGCTTGGGTCTCGAGGTTGG 3314
Qy 81 ThrProAlaSerCysHisleuPProleuArg-GlnSerAlaIleuHisProSerProSerle 100
Db 3315 ACCCCAGCTTCTTGGCCACCTTCCCTCCGGGCGATCAGCTCTCCATCCATCCCTCTT 3374
Qy 100 UllleTyrgluSerllleGlySerValCysValThr-ThrHisProTyArgCysProSerA 120
Db 3375 AATCTATGAATCTATAGGCTGGGTGTGTATACACACACCCCTATCGTTCCTTCA 3434
Qy 120 snThrglnHisTyRHisTTPleuArgProAsnSerGlnleuSerGlnIleArgPheThrI 140
Db 3435 ATACTCAGCATTCACATGGTGTAGGCCAAATTCAGAGCTTCTCAATACAGATTACAA 3494
Qy 140 lserIlePheIleAsnGlyGluThrSerProSerHis 152
Db 3495 TCTCCATTTCATTAAAGGGGAAACATCCCGAGCCAC 3532
RESULT 11
AA233583 ID AA233583 standard; cDNA; 2192 BP.
XX AC AA233583;
XX 08-DEC-1999 (first entry)
XX DE Human breast tumour-associated EST 43.
XX KM Expressed sequence tag; EST; human; breast; cancer; cyrostatic;
XX KM medicaments; gene therapy; treatment; fat metabolism; ss.
XX OS Homo sapiens.
XX PN DE19813835-A1.
XX PD 23-SEP-1999.
XX PF 20-MAR-1998; 98DE-01013835.
XX PR 20-MAR-1998; 98DE-01013835.
XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX PT Specht T, Hinemann B, Schmitt A, Pilsarsky C, Dahl E, Rosenthal A;
XX DR WPI; 1999-528979/45.
XX DR P-PSDB; AAY48501, AAY48502.
XX PT Human nucleic acid sequences and protein products from normal breast
XX PT tissue, useful for breast cancer therapy.
XX PS Claim 1a; 131; 206bp; German.

CC This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytoskeletal activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated with
CC fat metabolism. AA33541-233610 represent expressed sequence tags
CC described in the method of the invention

XX
SQ Sequence 2192 BP; 457 A; 717 C; 588 G; 430 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,25e-54	Length:	2192
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x AA33583 (1-2192)

QY 1 MepProHisLeuTYrIleAspGlyValPheProIleGlnLeuValArgGlnLeuHis 20
Db 1198 ATGCCAATCTGTATATAGATGGGCTTTTCCATACAGCTGTCTGTATTAACAGCAT 1257
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGly 40
Db 1258 GAAACCTCCGCGCTGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1217
QY 41 ThrGlyThrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaPheMetGlnThr 60
Db 1318 TGGGGCTGCT 1377
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
Db 1378 GGATCACCCTAACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1437
QY 81 ThrProHisLeuSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSer 100
Db 1438 ACCCAGCTCTTGGCAGCT 1497
QY 100 uileTYrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProser 120
Db 1498 AATTAATGAATCTATAGCTCTGCTGTGTATACACACACACACACACACACACACACAC 1557
QY 120 snTrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 1558 ATACTCAGCATATACATGTTGAGGCCAATTAGAGCTTCTCAATCAGATTACAA 1617
QY 140 IeserIlePheIleAsnGlyGlnThrSerProserHis 152
Db 1618 TCTCCATTTTCATTACGGGAAACATCCCGAGCCAC 1655

RESULT 12

ID ADM19231 standard; cDNA; 2220 BP.

XX ADM19231;

DT 20-MAY-2004 (first entry)

XX Novel human channel/transporter gene #28.

XX ds; Gene: immunosuppressive; antiarthritic; antineumatic;
KW antiproliferative; cytoskeletal; cardiac; vasotropic; cerebroprotective;
KW neurotrophic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmological; gene therapy; channel/transporter protein;
KW rheumatoid arthritis; neoplasia; cardiac arrest; cerebrovascular disorder;
KW cerebral ischemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.

XX Homo sapiens.
OS
XX
XX MO200154472-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225577P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0225797P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232977P.
PR 14-SEP-2000; 2000US-0232988P.
PR 14-SEP-2000; 2000US-0232999P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000, 2000US-0235484P.
 PR 27-SEP-2000, 2000US-0235834P.
 PR 27-SEP-2000, 2000US-0235836P.
 PR 29-SEP-2000, 2000US-0236327P.
 PR 29-SEP-2000, 2000US-0236367P.
 PR 29-SEP-2000, 2000US-0236368P.
 PR 29-SEP-2000, 2000US-0236369P.
 PR 29-SEP-2000, 2000US-0236370P.
 PR 02-OCT-2000, 2000US-0236802P.
 PR 02-OCT-2000, 2000US-0237037P.
 PR 02-OCT-2000, 2000US-0237038P.
 PR 02-OCT-2000, 2000US-0237039P.
 PR 02-OCT-2000, 2000US-0237040P.
 PR 13-OCT-2000, 2000US-0239935P.
 PR 13-OCT-2000, 2000US-0239937P.
 PR 20-OCT-2000, 2000US-0240960P.
 PR 20-OCT-2000, 2000US-0241221P.
 PR 20-OCT-2000, 2000US-0241785P.
 PR 20-OCT-2000, 2000US-0241786P.
 PR 20-OCT-2000, 2000US-0241787P.
 PR 20-OCT-2000, 2000US-0241808P.
 PR 20-OCT-2000, 2000US-0241809P.
 PR 01-NOV-2000, 2000US-0241826P.
 PR 08-NOV-2000, 2000US-0246174P.
 PR 08-NOV-2000, 2000US-0246474P.
 PR 08-NOV-2000, 2000US-0246475P.
 PR 08-NOV-2000, 2000US-0246476P.
 PR 08-NOV-2000, 2000US-0246477P.
 PR 08-NOV-2000, 2000US-0246478P.
 PR 08-NOV-2000, 2000US-0246523P.
 PR 08-NOV-2000, 2000US-0246524P.
 PR 08-NOV-2000, 2000US-0246525P.
 PR 08-NOV-2000, 2000US-0246526P.
 PR 08-NOV-2000, 2000US-0246527P.
 PR 08-NOV-2000, 2000US-0246528P.
 PR 08-NOV-2000, 2000US-0246532P.
 PR 08-NOV-2000, 2000US-0246609P.
 PR 08-NOV-2000, 2000US-0246610P.
 PR 08-NOV-2000, 2000US-0246611P.
 PR 08-NOV-2000, 2000US-0246613P.
 PR 17-NOV-2000, 2000US-0249207P.
 PR 17-NOV-2000, 2000US-0249208P.
 PR 17-NOV-2000, 2000US-0249209P.
 PR 17-NOV-2000, 2000US-0249210P.
 PR 17-NOV-2000, 2000US-0249211P.
 PR 17-NOV-2000, 2000US-0249212P.
 PR 17-NOV-2000, 2000US-0249213P.
 PR 17-NOV-2000, 2000US-0249214P.
 PR 17-NOV-2000, 2000US-0249215P.
 PR 17-NOV-2000, 2000US-0249216P.
 PR 17-NOV-2000, 2000US-0249217P.
 PR 17-NOV-2000, 2000US-0249218P.
 PR 17-NOV-2000, 2000US-0249244P.
 PR 17-NOV-2000, 2000US-0249245P.
 PR 17-NOV-2000, 2000US-0249246P.
 PR 17-NOV-2000, 2000US-0249265P.
 PR 17-NOV-2000, 2000US-0249265P.
 PR 17-NOV-2000, 2000US-0249297P.
 PR 17-NOV-2000, 2000US-0249300P.
 PR 01-DEC-2000, 2000US-0250160P.
 PR 01-DEC-2000, 2000US-0250391P.
 PR 05-DEC-2000, 2000US-0251030P.
 PR 05-DEC-2000, 2000US-0251888P.
 PR 05-DEC-2000, 2000US-0251889P.
 PR 05-DEC-2000, 2000US-0251479P.
 PR 08-DEC-2000, 2000US-0251856P.
 PR 08-DEC-2000, 2000US-0251868P.
 PR 08-DEC-2000, 2000US-0251869P.
 PR 08-DEC-2000, 2000US-0251989P.
 PR 08-DEC-2000, 2000US-0251990P.
 PR 11-DEC-2000, 2000US-0254097P.
 PR 05-JAN-2001, 2001US-0259678P.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI, 2001-476159/51.
 DR P-PSDB; ADM19710.
 XX
 PT Isolated nucleic acid molecule encoding a channel/transporter protein is
 PR used in preventing, treating or ameliorating a medical condition.
 XX
 PS Claim 1, SEQ ID NO 38; 809gp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule encoding a
 CC channel/transporter protein or sequences at least 95% identical to a
 CC these. The nucleic acids and proteins encoded by them are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. The antibodies to the proteins can also be used
 CC in alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. This sequence corresponds to a gene of the
 CC invention.
 XX
 SQ Sequence 2220 BP; 461 A; 718 C; 597 G; 429 T; 0 U; 15 Other;
 XX
 Alignment Scores:
 Pred. No.: 1,27e-54 Length: 2220
 Score: 811.00 Matches: 149
 Percent Similarity: 97.40% Conservative: 1
 Best Local Similarity: 96.75% Mismatches: 2
 Query Match: 95.64% Indels: 0
 DB: 5 Gaps: 0
 US-09-977-418-8 (1-152) x ADM19231 (1-2220)
 QY 1 MetProHisLeuYrTleAepGlyValPheProIleGlnLeuValArgGlyLysLeuHis 20
 DB 1252 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGGTGTGATAACACGCAAT 1311
 QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysIleTyrPgiY 40
 DB 1312 GAATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1371
 QY 41 TyrGlyTyrPsePheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 DB 1372 TGGGGCTGTCT 1431
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 1432 GGATCACCCTAACAGAGTGGAGCCAGGCGATGATGGAGCTTTGGAGTCTCTGAGGTTGG 1491
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 1492 ACCCCAGCTTCTTGCCACACCTTCCCTCCGAGGAGTCAAGCTCCATCCATCCCTCTTT 1551
 QY 100 uileTyrGlnSerIleGlySerAlaCysValThr-ThrHisProTyrArgCysProSerA 120
 DB 1552 AATCTATGAATCTATAGGCTGGGTGTGTGAACACACACCCCTATCTGTCTCTCA 1611
 QY 120 snThrGlnHisTyrHisTyrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140

Db 1612 ATACTGACATTCATTCGTTGAGGCCAAATTCAGAGCTTCTCAATCAGATTACAA 1671
Qy 140 |leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1672 TCTCCATTTTCATTACGGGGAACATCCCGAGCCAC 1709

RESULT 13
ID AAK94832 standard; cDNA; 2701 BP.
XX
AC AAK94832;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3983.
XX
KM Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
P-PSDB; AAM93873.
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 8; SEQ ID NO 3983; 1380bp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2701 BP; 557 A; 867 C; 771 G; 506 T; 0 U; 0 Other;

Alignment Scores:
Pred. NO: 1.58e-54 Length: 2701
Score: 811.00 Matches: 149
Percent Similarity: 97.40% Conservative: 1
Best Local Similarity: 96.75% Mismatches: 2
Query Match: 95.64% Indels: 2
DB: 4 Gaps: 0

US-09-977-418-8 (1-152) x AAK94832 (1-2701)

Qy 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1766 ATGCCACATCTGTATATAGATGGGCTTTTCATATACAGCTGGTTCGTATTAACATGCAAT 1825
Qy 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrPglyTyrPgly 40

Db 1826 GAAACTCCTGCCTGCTCTGCGCTGCTGGGGCTCCAGGCAAGGCGCACCTGGGGTGGGGG 1885
Qy 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db 1886 TGGGGCTGGTCTCTTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGCAAGCA 1945
Qy 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 1946 GGATCACCCTAACAGAGATGAGAGCCAGGCGATGATGGGCTTGGGGCTCCGAGGTTGG 2005
Qy 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 2006 ACCCCAGCTTCTTGGCAGCTTCCCTCCGGGAGAGCAGCTCTCATTCATCCCTCTTT 2065
Qy 100 vIleTyrGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
Db 2066 AATCTATGATATATAGGCTCGGTGTGTGTAACACACACACCCCTATGCTGTCTTCA 2125
Qy 120 sTrThrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 2126 ATACTGACATTCATTCGTTGAGGCCAAATTCAGACCTTCTCAATCAGATTACAA 2185
Qy 140 |leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 2186 TCTCCATTTTCATTACGGGGAACATCCCGAGCCAC 2223

RESULT 14
ID ADJ31950 standard; cDNA; 2701 BP.
XX
AC ADJ31950;
XX
DT 20-MAY-2004 (first entry)
XX
DE Full length human cDNA clone SeqID 3983.
XX
KM human; medicine; signal transduction; glycoprotein; transcription;
KM oligo-capping method; ss; gene.
XX
OS Homo sapiens.
XX
PN EP1396543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2004-204755/20.
XX
P-PSDB; ADJ31951.
PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
PS Example 1; SEQ ID NO 3983; 1340bp; English.
XX
CC This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'

CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.

SO Sequence 2701 BP; 557 A; 867 C; 771 G; 506 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,58e-54	Length:	2701
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	12	Gaps:	0

US-09-977-418-8 (1-152) x ADL1950 (1-2701)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
 Db 1766 ATCCACATCTGATATAGATGGCGTTTCCAAATACACCTGCTGATATAACGCAT 1825
 QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTrpGly 40
 Db 1826 GAAATCTCCGCTCTCCGCTGCGCTGCGGCTCCAGGCAAGCCACGCGGGGTTGGGG 1885
 QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 1886 TGGGGCTGCTCTCTCCCTCCACAGCGCTGCTTCTGGGGCTGCTCCATGACAGCA 1945
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValLeuGlyProArgGlyTrp 80
 Db 1946 GGATCACTCAACAGAAATGGAAGCCAGGCGCATGATGAGGGGCTTGGGCTCTGAGAGTTGG 2005
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 2006 ACCCGAGCTTCTTCCACCTTCCCTCCGCGGAGCTGCTCCATCCATCCCTCTTT 2065
 QY 100 ValTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
 Db 2066 AATCTATGATCAATCAATGAGCTCGCTGCTGTTACACACACACCCCTATGCTTCTTCAA 2125
 QY 120 snThrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnLeuArgPheThr 140
 Db 2126 ATTCCTAGCATTCACCATTCGTTGAGGCGCAAAATTCAGAGCTTCTCAAATTCAGATTACAA 2185
 QY 140 IseSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 2186 TCTCCATTTTCATTACGCGGAAACATCCCGAGGCAC 2223
 RESULT 15
 ADL29984
 ID ADL29984 standard; cDNA; 2783 BP.
 AC ADL29984;
 XX 18-DEC-2003 (first entry)
 DT
 XX Human novel cDNA sequence, SEQ ID NO:66.
 DE
 XX Human; diagnostic; drug screening; forensics; gene mapping;
 XX neurodegenerative assessment; Parkinson's disease; Alzheimer's disease;
 XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 XX ulcers; osteoporosis; autoimmune disease; cancer;
 XX molecular weight marker; food supplement; antiparkinsonian; nootropic;
 XX neuroprotective; antianemic; anticoagulant; thrombolytic; vihenary;
 XX anticancer; osteopathic; immunosuppressive; antiinflammatory; cyostatic;
 XX gene therapy; chromosome 1p36.11-36.2; gene; ss.
 OS
 XX Homo sapiens.
 OS
 PN MO2003029271-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou F, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;
 DR WPI; 2003-371981/35.
 DR P-PSDB; ADL0955.
 DR
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 PS Claim 1; SEQ ID NO 66; 1185bp; English.
 XX
 CC The invention relates to 971 novel human cDNA sequences (ADL29984-
 CC ADL30889) and the polypeptides they encode (ADL30890-ADL31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC coding sequences corresponding to the cDNA sequences of the invention
 CC (ADL31861-ADL32627) and the polypeptides encoded by the cDNAs (ADL32628
 CC ADL33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human cDNA sequence of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 2783 BP; 586 A; 880 C; 781 G; 536 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.64e-54	Length:	2783
Score:	811.00 <td>Matches:</td> <td>149</td>	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	10	Gaps:	0

US-09-977-418-8 (1-152) x ADL29984 (1-2783)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
 Db 1833 ATCCACATCTGATATAGATGGCGTTTCCAAATACACCTGCTGATATAACGCAT 1892
 QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTrpGly 40
 Db 1893 GAAATCTCCGCTCTCCGCTGCGCTGCGGCTCCAGGCAAGCCACGCGGGGTTGGGG 1952
 QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60

Db	1953	TGGGGCTGTCCTTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGACACA	2012
Qy	61	GlySer-ProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr	80
Db	2013	GGATCACCTTAACAGAGATGAGAACCGAGGCGATGATGGGCGCTTGGGCTCCGAGTTGG	2072
Qy	81	ThrProAlaSerCysHisLeuProLeuArg-GlnSerIleLeuHisProSerProSerIe	100
Db	2073	ACCCGAGTTCTTGGCACCTTCCCTCCGGGAGTCACTTCCATTCATCCCTCTTT	2132
Qy	100	uIeTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA	120
Db	2133	AATCATATAACTATAGGCTCGGTGTGTGTAAACACACACACCCCTATGGTGTCTTCAA	2192
Qy	120	snThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI	140
Db	2193	ATATCACACATTACCATGTTGTGGGCCAAATTGAGCTTCTCAATCATGATTTTCAA	2252
Qy	140	IeSerIlePheIleHisGlnGlyIuThrSerProSerHis	152
Db	2253	TCTCCATTTTCATTACGGGGAAACATCCCGAGGCAC	2290

Search completed: October 24, 2004, 16:12:58
Job time : 435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:05:43 ; Search time 424 Seconds

(without alignments)
1836.877 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848

Sequence: 1 MPHLYIDGVPPILQVREKLN.....LSQRFTSIFINGETSPSH 152

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3407233 segs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6614466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q=/cgnt2_1/USPFO.spool/US09977418/runat.22102004.115949.1078/app.query.fasta.1.327
-DB=Published Applications NA -QPM=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCT=0 -IOGEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=bitsune2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO MMAP -LARGECQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications_NA:*
1: /cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgnt2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgnt2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgnt2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgnt2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgnt2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgnt2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgnt2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgnt2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
13: /cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
18: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
19: /cgnt2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgnt2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cgnt2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	846	99.8	1987	10	US-09-977-418-7
2	846	99.8	1987	10	US-09-977-033A-7
3	846	99.8	1987	10	US-09-977-751C-7
4	846	99.8	1987	10	US-09-977-639A-7
5	846	99.8	1987	11	US-09-977-639B-7
6	802	94.6	2410	15	US-10-104-047-972
7	644	75.9	400	10	US-09-918-995-447
8	512	60.4	305	10	US-09-977-418-75
9	512	60.4	305	10	US-09-977-033A-75
10	512	60.4	305	10	US-09-977-751C-75
11	512	60.4	305	10	US-09-977-639A-75
12	512	60.4	305	11	US-09-977-639B-75
13	106.5	12.6	1233	17	US-10-437-963-55452
14	104.5	12.3	9025608	15	US-10-437-963-48431
15	104	12.3	9025608	15	US-10-437-963-48431
16	103.5	12.2	21010	10	US-09-764-891-8576
17	103.5	12.2	21024	10	US-09-764-891-8577
18	103	12.1	1701	15	US-10-029-386-20567
19	103	12.1	2888	16	US-10-108-280A-1403
20	103	12.1	2854	16	US-10-220-891-9
21	103	12.1	4810	16	US-10-188-646-11
22	102.5	12.1	5698	16	US-10-420-191-1
23	102	12.0	131576	13	US-10-087-192-1564
24	101	11.9	341	9	US-09-867-701-3673
25	101	11.9	84410	17	US-10-322-281-747
26	99	11.7	5493	15	US-10-408-501-5
27	98	11.6	1385	17	US-10-437-963-82042
28	98	11.6	2483	15	US-10-104-047-1600
29	97.5	11.5	31652	13	US-10-087-192-406
30	97.5	11.5	60430	13	US-10-087-192-1732
31	97.5	11.5	77478	13	US-10-087-192-1378
32	97	11.4	1330	15	US-10-322-813-17
33	97	11.4	1330	15	US-10-322-813-17
34	97	11.4	1330	17	US-10-729-668-17
35	97	11.4	1313	14	US-10-717-049-17
36	97	11.4	2000	9	US-10-116-802-31
37	97	11.4	2000	9	US-09-969-708-276
38	97	11.4	2000	9	US-09-919-497-10
39	97	11.4	2000	9	US-09-880-107-3774
40	97	11.4	2000	10	US-09-960-706-1082
41	97	11.4	2000	16	US-10-305-720-1291
42	97	11.4	2000	16	US-10-240-445-1545
43	97	11.4	2015	17	US-10-775-169-85
44	97	11.4	2047	17	US-10-341-434-92
45	97	11.4	2616	10	US-10-115-635-46
					US-09-971-429B-41

ALIGNMENTS

RESULT 1
US-09-977-418-7
Sequence 7, Application US/09977418
Publication No. US20030027158A1
GENERAL INFORMATION:
APPLICANT: Shimkete et al
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded by
FILE REFERENCE: 15966-552
CURRENT APPLICATION NUMBER: US/09/977,418
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)..(1446)
NAME/KEY: variation
LOCATION: (1)..(1981)
OTHER INFORMATION: where n can be any nucleotide
US-09-977-418-7

Alignment Scores:
Pred. No.: 2,53e-87 Length: 1987
Score: 846.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-418-7 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluYsLeuHis 20
DB 991 ATGCCACATCTGATATAGATGGGTTTTCATACAGCTGGTGTGAAAACTGCAT 1050
QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyValAlaYsTrpGlyTPGly 40
DB 1051 GAAACTCTGCGCTCTGCGCTGCGGCGCTCCAGGCAAGCCAAAGTGGGTTGGGGG 1110
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1111 TGGGGTGTGCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGCTCCATGCAGACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValLeuGlyProArgGlyTP 80
DB 1171 GGATCACCCTAAGAGATGAGCCAGGCGGATGGAGGCTTGGGCTCTCGAGGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGACAGCTCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTATGAATCTATAGCTCGGTGTGTGTACACACACCCCTATCGTGTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTrpLeuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTCAGCATTTACCATTTGAGGCGCAAAATTCAGAGCTTCTCAATTCAGATTTCATC 1410
QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
DB 1411 TCCATTTTCAATTACGGGMAACATCCCGAGCCAC 1446

RESULT 2
US-09-977-033A-7
Sequence 7, Application US/0977033A
Publication No. US2003008255A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
APPLICANT: Verneil, Corine
TITLE OF INVENTION: NO. US2003008255A1el nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: protein-like and human protein PRO288-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977, 033A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137, 322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189, 810

PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191, 158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193, 086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201, 388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584, 411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)..(1446)
NAME/KEY: variation
LOCATION: (1148)
OTHER INFORMATION: wherein n is an a or t or c or g
NAME/KEY: variation
LOCATION: (1498)
OTHER INFORMATION: wherein n is an a or t or c or g
NAME/KEY: variation
LOCATION: (1925)
OTHER INFORMATION: wherein n is an a or t or c or g
NAME/KEY: variation
LOCATION: (1943)
OTHER INFORMATION: wherein n is an a or t or c or g
NAME/KEY: variation
LOCATION: (1969)..(1972)
OTHER INFORMATION: wherein n is an a or t or c or g
US-09-977-033A-7

Alignment Scores:
Pred. No.: 2,53e-87 Length: 1987
Score: 846.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-033A-7 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluYsLeuHis 20
DB 991 ATGCCACATCTGATATAGATGGGTTTTCATACAGCTGGTGTGAAAACTGCAT 1050
QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyValAlaYsTrpGlyTPGly 40
DB 1051 GAAACTCTGCGCTCTGCGCTGCGGCGCTCCAGGCAAGCCAAAGTGGGTTGGGGG 1110
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1111 TGGGGTGTGCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGCTCCATGCAGACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValLeuGlyProArgGlyTP 80
DB 1171 GGATCACCCTAAGAGATGAGCCAGGCGGATGGAGGCTTGGGCTCTCGAGGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGACAGCTCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTATGAATCTATAGCTCGGTGTGTGTACACACACCCCTATCGTGTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTrpLeuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTCAGCATTTACCATTTGAGGCGCAAAATTCAGAGCTTCTCAATTCAGATTTCATC 1410

Alignment Scores:	
Pred. No.:	2,53e-87
Score:	846.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	99.76%
DB:	10
Length:	1987
Matches:	152
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

```

QY      MetProH1sLeuYrY11aAspG1yVa1PhePro11eG1nLeuVa1aG1u1ySLeuH1s 20
Db      991 ATGCCACATCTGTATATATGATGGGGCTTTTCCATACAGCTGGTTCGTGAAAACTGCAT 1050
QY      21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLySa1a1yS1rTpG1y 40
Db      1051 GAACCTCGCGCTCTCGCTCGCTCGCTGGGGCTCCAGGCAAGGCCAAGGGGCTGGGGG 1110
QY      41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
Db      1111 TGGGGCTGTGCTCTTCCTCCCTCCCAAGGCTGTGTCTGTGGGGCTGCTCCCATGACACA 1170
QY      61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrp 80
Db      1171 GGATACCTTAACAGAGATGAGAGCCAGGAGCATGATGGGGCTTGGGCTCTCGAGTGGG 1230
QY      81 ThrProAlaSerCysH1sLeuProLeuArgGlnSerAlaLeuH1sProSerProSerLeu 100
Db      1231 ACCCAGCTTCGTGCCACTTCCTCCCTCCGGCAGCTTCAGCTTCACCTCCCTCTCTTAA 1290
QY      101 IleTrpGluSer11eGlySerValCysValThrThrH1sProTyArgCysProSerAsn 120
Db      1291 ATCTATGATCTATATGGCTCGCTGGTGTGTAAACAACAACCCCTATCGTGTCTCTCAAT 1350
QY      121 ThrGlnH1sYrH1s1rTpLeuArgProAsnSerGluLeuSerGln11eArgPheThr11e 140
Db      1351 ACTCGACATTACCATTTGGTTGAGGCCAAATTACAGCTTTCCTCAATCAGATTTCATTC 1410
QY      141 Ser11ePhe11eAsnGlyGluThrSerProSerH1s 152
Db      1411 TCCATTTCATTAAACGGGAAACATCCCGAGCCAC 1446

RESULT 5
US-09-977-819B-7
; Sequence 7, Application US/09977819B
; Publication No. US20040002134A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Fernandes, Elma
; APPLICANT: Herrman, John
; APPLICANT: Vernet, Corine

```

```

; SEQ ID NO 1
; LENGTH: 1987

```

ORGANISM: *Homo sapiens*

FEATURE: NAME/REV: CDD

NAME/NET: CDS
LOCATION: (991)..(1446

FEATURE:

```

; NAME/KEY: misc feature
; LOCATION: (1148) ... (1148)

```

OTHER INFORMATION: "U" =
REACTIVE.

```

;
FEATURE:
NAME/KEY: misc_feature

```

LOCATION: (1498) ... (1498)
OTHER INFORMATION: "U" =

FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (1925)...(1925)

```

OTHER INFORMATION: "N" =
FEBRUARY.

```

;
FEATURES:
NAME/KEY: misc_feature

```

LOCATION: (1943) ... (1943)
OTHER INFORMATION: "n" =

FEATURE:

```

; NAME/KEY:  MISC Feature
; LOCATION:  (1969) ... (1972
;

```

OTHER INFORMATION: "U" =
FEATURE.

```
NAME/KEY: misc_feature
```

```

; LOCATION: (53)... (53)
; OTHER INFORMATION: "Xaa"

```

US-09-977-819B-7

Alignment Scores:

Pred. No.:	2.53
Score:	846

Percent Similarity: 100.

Best Local Similarity:	100.
Query Match:	99.7

DB: 11

US-09-977-418-8 (1-152) x US

1 MetProH1stLeuTy

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db 991 ATGCCACATCTGTA

QY 21 GlutHrProAlava

Db 1051 GAAACTCCTGCCGT

Or 41 TmG1vTmSeph

Db 1111 TGAGGCTGGCTTCTCCCTCCACAGGCTGTGTCNTGGGGCTGCTCCCATGACAGACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1171 GGATCACTTAACAAGATGAGAGAGCCAGGCGATGATGGGGCTTTGGGTCTCCAGAGTTGG 1230
 QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 1231 ACCCGAGCTTCTCCACCTTCCCTCCCGGAGTACAGTCTCCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 Db 1291 ATCTATGATCTTAAGGCTCGGTGTGTGTATACACACACCCCTATCGTTGTCTTCAAT 1350
 QY 121 ThrGlnHisTyrHisTyrPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
 Db 1351 ACTCAGCATTCACCTGGTTGGTGAAGCCCAATTCAGAGCTTCTCTCAATCAGATTACATC 1410
 QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 1411 TCCATTTTCATTAACGGGGAACAATCCCGAGCCAC 1446

RESULT 6
 US-10-104-047-972
 ; Sequence 972, Application US/10104047
 ; Publication No. US20030236392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
 ; FILE REFERENCE: H1-A0105
 ; CURRENT APPLICATION NUMBER: US/10/104,047
 ; CURRENT FILING DATE: 2002-03-25
 ; PRIOR APPLICATION NUMBER:
 ; PRIOR FILING DATE:
 ; NUMBER OF SEQ ID NOS: 4096
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 972
 ; LENGTH: 2410
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-104-047-972

Alignment Scores:
 Pred. No.: 3,72e-82 Length: 2410
 Score: 802.00 Matches: 148
 Percent Similarity: 96.75% Conservative: 1
 Best Local Similarity: 94.10% Mismatches: 3
 Query Match: 94.58% Indels: 2
 DB: 15 Gaps: 0

US-09-977-418-8 (1-152) x US-10-104-047-972 (1-2410)

QY 1 MetProHisLeuTyrTyrIleAspGlyValPheProIleGlnLeuValArgGlnIleuHis 20
 Db 1478 ATGCCAATCTGTATAGATGAGGAGTTTTCATTAAGCTGTGTCGTATTAACAGCA 1537
 QY 21 GlnThrProAlaValLeuArgLeuGlyProProGlyLysAlaIleArgTyrGly 40
 Db 1538 GAAACTCTCGCGCTCGCGCTGCTGGGCGCTCCAGCAAGCCCAAGCTGGGTTGGG 1597
 QY 41 TyrGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaIleProMetGlnThr 60
 Db 1598 TGGGCTGTGTCTTCTCTCCCTCCACAGGCTGTGTTCTTGGGGCTGCTCCATGACAGACA 1657
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1658 GGATCACTTAACAAGATGAGAGAGCCAGGCGATGATGGGGCTTGGGTCTCTCAAGTTGG 1717
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100
 Db 1718 ACCCGAGCTTCTCCACCTTCCCTCCCGGAGTACAGTCTTCATCCATCCCTCTTT 1777
 QY 100 uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSer 120

Db 1778 AATCTATGATCTTAAGGCTCGGTGTGTGTATACACACACCCCTATCGTTGCTTCA 1837
 QY 120 snThrGlnHisTyrHisTyrPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
 Db 1838 ATACTCAGATTACCAATGGTTGAGGCAAAATTCAGAGCTTCTCAATCAGATTACAA 1897
 QY 140 IeSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 1898 TCTCATTTTCATTAACGGGGAACAATCCCGAGCCAC 1935

RESULT 7
 US-09-918-995-4147
 ; Sequence 4147, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HVBeg, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4147
 ; LENGTH: 400
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-918-995-4147

Alignment Scores:
 Pred. No.: 6.37e-65 Length: 400
 Score: 644.00 Matches: 118
 Percent Similarity: 97.52% Conservative: 0
 Best Local Similarity: 97.52% Mismatches: 1
 Query Match: 75.94% Indels: 2
 DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-918-995-4147 (1-400)

QY 34 LysAlaIleTyrPLeuArgGlyTyrPLeuArgTyrPheSerPheSerLeuProGlnAlaCysVal** 53
 Db 3 AAGCCCAAGTGGGCTGGGGGCTGGGCTGCTCTCTCCCTCCACAGGCTGTGTTCTT 62
 QY 54 GlyAlaIleProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyMetAspGly 73
 Db 63 GGGGCTGTCTCCCATCAGACAGATCACTTAACAGATGAGAGCCAGGCGATGGATGG 122
 QY 74 AlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuProLeuArg-GlnSerAl 93
 Db 123 GCTTTGGGCTCTCGAGTTGAGACCCAGCTTCTTCCCTCCAGGCGATGAC 182
 QY 93 AlaLeuHisProSerProSerLeuIleTyrGlnSerIleGlySerValCysValThr-Thr 113
 Db 183 TCTCATCATCCCTCTTATCTATGAAATCTATAGAGTGGGTGTGTATACACACAC 242
 QY 113 IeProTyrArgCysProSerAsnThrGlnHisTyrHisTyrPLeuArgProAsnSerGln 133
 Db 243 ACCCGATGCTGTCTCTCAAAATCTCAAGCATTCATGATGGTGAAGCAAAATTCAGAGC 302
 QY 133 eUeSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 303 TTTTCAATCAGATTTCATTCATCTTCATTTCAATGAAGGGAACAATCCCGAGCCAC 361

RESULT 8
 US-09-977-418-75
 ; Sequence 75, Application US/09977418
 ; Publication No. US20030027158A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shinkets et al
 ; TITLE OF INVENTION: NO. US20030027158A1e1 polynucleotides and polypeptides encoded c)

FILE REFERENCE: 15966-552
CURRENT APPLICATION NUMBER: US/09/977,418
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 75
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
US-09-977-418-75

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-418-75 (1-305)

QY 1 MetProHisLeuTyrlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 1 ATGCCACATCTGATATAGATGGGCTTTTCCAAATACAGCTGTTGCTGATAAAGTGCAT 60
QY 21 GlnThrProAlaValLeuArgLeuGlnLysProProGlyLysAlaLysTrpGlyTyrGly 40
DB 61 GAAACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 120
QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTCCTCTCTCCCT-CCCGAGGCTGTGTTCTTGGGGCTGCTCCATGCAGACA 179
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACCCTAAGACAGATGGAGCCAGGCGCATGGATGGGCTTGGGCTCTGAGGTTGG 239
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCCAGCTTCTTCCACCTTCCCTCCGGGCGAGTCAGCTCCATCCATCCCTCTTT 299
QY 100 uile 101
DB 300 AATC 303
RESULT 9
US-09-977-033A-75
Sequence 75, Application US/09977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herman, John
TITLE OF INVENTION: No. US20030082554A1el nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977,033A
CURRENT FILING DATE: 2001-10-15

PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 75
LENGTH: 305
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-033A-75

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-033A-75 (1-305)

QY 1 MetProHisLeuTyrlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 1 ATGCCACATCTGATATAGATGGGCTTTTCCAAATACAGCTGTTGCTGATAAAGTGCAT 60
QY 21 GlnThrProAlaValLeuArgLeuGlnLysProProGlyLysAlaLysTrpGlyTyrGly 40
DB 61 GAAACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 120
QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTCCTCTCTCCCT-CCCGAGGCTGTGTTCTTGGGGCTGCTCCATGCAGACA 179
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACCCTAAGACAGATGGAGCCAGGCGCATGGATGGGCTTGGGCTCTGAGGTTGG 239
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCCAGCTTCTTCCACCTTCCCTCCGGGCGAGTCAGCTCCATCCATCCCTCTTT 299
QY 100 uile 101
DB 300 AATC 303
RESULT 10
US-09-977-751C-75
Sequence 75, Application US/0997751C
Publication No. US20030134430A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herman, John
TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhabdit
TITLE OF INVENTION: Polypeptides.
FILE REFERENCE: 15966-552 CON S-40
CURRENT APPLICATION NUMBER: US/09/977,751C
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322

PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/564,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 75
LENGTH: 305
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-819B-75

Alignment Scores:
Pred. No.: 7.15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 11 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-819B-75 (1-305)

QY 1 Me-ProHisLeuTyrIleAspGlyValPheProIleGluLeuValArgGluLysLeuHis 20
DB 1 ATCCCATCTGTATATAGATGGGTTTCCAAATCAGCTGTGCTATTAACATGCAAT 60

QY 21 GluThrProAlaValLeuArgLeuGluGlyProProGlyLysAlaLysTrpGlyTyrGly 40
DB 61 GAACACTCTGCTCCCTCGCGCTCGCGGAGCCCTCCGAGGCAAGCCAGTGGAGTGGGG 120

QY 41 TyrGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTGCTCTCCCT-CCCGAGGCTGTGTTGGGGCTGCTCCCATGACAGCA 179

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACTTAACAGATGGAAGCCAGGCGATGATGGAGGCTTTGGGCTCTGAGGTTGG 239

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCGAGCTTCTTCCACCTTCCCTCCGCGGAGTCAAGCTTCCATCCATCCCTCTTT 299

QY 100 uile 101
DB 300 AATC 303

RESULT 13
US-10-437-963-55452
Sequence 55452, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 55452
LENGTH: 1213
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_57458C.1
US-10-437-963-55452

Alignment Scores:
Pred. No.: 0.0163 Length: 1213
Score: 106.50 Matches: 42
Percent Similarity: 35.58% Conservative: 16
Best Local Similarity: 25.77% Mismatches: 68
Query Match: 12.56% Indels: 37
DB: 17 Gaps: 7

US-09-977-418-8 (1-152) x US-10-437-963-55452 (1-1213)

QY 10 PheProIleGluLeuValArgGluLysLeuHisGluThr---ProAlaValLeuArgLeu 28
DB 9 TTTCACACCCCAACCAATACCGAGCGGAGCTTGACCGCGCGGAGCAGCCGCAATC 68

QY 29 LeuGlyProProGlyLysAlaLysTyrGlyTyrGly----- 42
DB 69 GCGGAGCAAGCGGAGCGGAGGATGGAGTGGGATCGAGCGGAGCGGAGCAAGCCGCTGC 128

QY 43 -----TyrSerPheSerLeuProGlnAla 50
DB 129 GCGCGATCGCGGCGTCTGAGAGAGCTCGAGCGGCTGGCGCAAGCAGCGCCCGCGCC 188

QY 51 CysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGly 70
DB 189 CGGCGATGAGAGTGGGCTCTTGGCCCGCGCTGCTCAAGCTCTCCCTCTTGGGCT 248

QY 71 MetAspGly-----AlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuPro 88
DB 249 GCGTGGCATCGGCTTCAAGTTCGCGAGG---TGGAGTACGTGCGCAAGCAAGCCCC 305

QY 89 LeuArgGlnSerAlaLeuHisProSerProSerLeuLysTrpGluSerIleGlySerVal 108
DB 306 GCCCGCACCCCGCTCTAGTCTTCGCTCCCTGCTCTCCAGAGATGAGTGGCGCTTA 365

QY 109 CysValThrThrHis-----ProTyrArgCysProSerAsn 120
DB 366 TTAATGTTGCTACTACTCTCCGTTTCAGGTTGATGATTGCTGAGGAGCATCCAAATGCA 425

QY 121 ThrGlnHis---TyrHisTrpLeu-----ArgProAsnSerGluLeu 133
DB 426 TCTCAAAATTCCTGAGATGTTGATGACATTGACATTGAGAGGACCGCATCAGCAAGCTG 485

QY 134 SerGlnIle 136
DB 486 GGAGCCATA 494

RESULT 14
US-10-437-963-48431
Sequence 48431, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 48431
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_51106C.1
US-10-437-963-48431

Alignment Scores:
Pred. No.: 0.0154 Length: 681
Score: 104.50 Matches: 40
Percent Similarity: 39.66% Conservative: 6
Best Local Similarity: 34.48% Mismatches: 47
Query Match: 12.32% Indels: 23
DB: 17 Gaps: 7

US-09-977-418-8 (1-152) x US-10-437-963-48431 (1-681)

QY 30 GlyProProlGlySerAlaLeuTyrGly-----TyrGlyTyrSerPhe 45
Db 191 GGGCCCTCCCGGAGG-----GGAGTCAGGCCCGAGGTAGGGTGGCGGCCA 235
QY 46 SerLeuProGlnAlaCysVal**GlyAlaAla---ProMetGlnThrGlySerProAsn 64
Db 236 CTCCTTCCCGCAAGAGCTAGTGGAGAGGCGTCCCGCGAGTCCACAGTGGCCCTCCACAG 235
QY 65 ArgAspGlySerGlnGlyMetAspGlyAlaLeuGly-----ProArgGly 79
Db 296 AGG-----GGAGTCAGGCCCGAGGTGGGGGGGTGGCCACTCTTCCCGAGAGAC 343
QY 80 TyrThrProAlaSer-----CysHisLeuProLeuArgGlnSerAlaLeuHisPro 96
Db 344 TGGATGAGAGAGTCTGGCCCTCCCATGCGACGTGGCCCTCCCGCAGGGGGGGTGGGGCCG 403
QY 97 ---SerProSerLeuLeuTyrGlnSerIleGlySerValCysValThrThrHisProTyr 115
Db 404 AGGTACACCGACGTAGCGTGTAGGTGCGCTCCGCTCCAGGTGGCCACTCATTTAT 463
QY 116 ArgCysProSerAsnThrGlnHisTyrHisTyrLeuArgProAsnSer 131
Db 464 GGGCAGCCGAGCGGTGGCGCCAGCCGCAATTAAAGGTGCGTGTCTCG 511

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10:156761
; Publication No. US20030113018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
```

```

; NAME/KEY: misc feature
; LOCATION: (418715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.93e+03 Length: 9025608
Score: 104.00 Matches: 32
Percent Similarity: 37.70% Conservative: 14
Best Local Similarity: 26.23% Mismatches: 38
Query Match: 12.28% Indels: 38
DB: 15 Gaps: 7

US-09-977-418-8 (1-152) x US-10-156-761-1 (1-9025608)

QY 32 ProGlyLysAlaLeuTyrGlyTyrGly----- 40
Db 4141353 CGAGGCAAGAGAGTGGGGCTGGTCGACGAGCTGTGAGAGAGGCCAGACCGACGG 4141294
QY 41 -----TyrGlyTyrSerPheSerLeuProGln-----AlaCysVal**Gly 54
Db 4141293 AGGCGCTGGCCTGGCCTTCCGACATGCGCGGACGTCACCGGTGGCCTGGCGCGCA 4141234
QY 55 AlaAlaPro-----MetGlnThrGlySerPro-----Asn 64
Db 4141233 AGGCGCCTCGCGCCTCGGCGCCAGGGCTGACCTGCGGGGGGCGCTGGAGTGGAGAGC 4141174
QY 65 ArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyrThrProAlaSer 84
Db 4141173 CGGCTGGGGCTCGTGGCTTCTCCGGGACCGGGCCGAGAGGGTGGGGCTTCACAG 4141114
QY 85 CysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeuLeuTyrGlnSer 104
Db 4141113 -----AGAGCGGCAAGCCGAGTGGCCCGCGC-----AGT 4141084
QY 105 IleGlySerValCysValThrThrHisProTyrArgCys---ProSerAsnThrGlnHis 123
Db 4141083 AGGCGCCCGGCGCGCGCGCGCGGACATCCCACTGAGGCCCGGCCCATGTCCCAAT 4141024
QY 124 TyrHis 125
Db 4141023 TGTCAAT 4141018

Search completed: October 24, 2004, 18:29:02
Job time : 1812 secs
```

